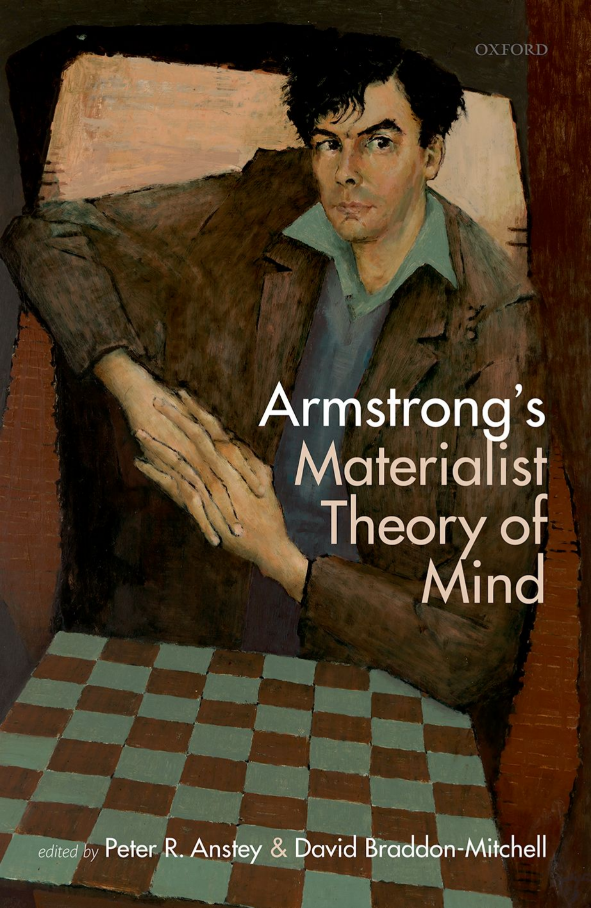

OXFORD

Armstrong's Materialist Theory of Mind

edited by Peter R. Anstey & David Braddon-Mitchell

Armstrong's Materialist Theory of Mind

Armstrong's Materialist Theory of Mind

Edited by

PETER R. ANSTEY AND
DAVID BRADDON-MITCHELL

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© the several contributors 2022

The moral rights of the authors have been asserted

First Edition published in 2022
Impression: 1

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by licence or under terms agreed with the appropriate reprographics rights organization. Enquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2021945310

ISBN 978–0–19–284372–2

DOI: 10.1093/oso/9780192843722.001.0001

Printed and bound in Great Britain by
Clays Ltd, Elcograf S.p.A.

Links to third party websites are provided by Oxford in good faith and for information only. Oxford disclaims any responsibility for the materials contained in any third party website referenced in this work.

Contents

Acknowledgements vii
Abbreviations ix
Contributors xi

Introduction 1
Peter R. Anstey and David Braddon-Mitchell

1. *A Materialist Theory of the Mind* in Context 9
Peter R. Anstey and David Braddon-Mitchell

2. Lewis and the Identity Theory 24
D. M. Armstrong

3. The Two Davids and Australian Materialism 29
A. R. J. Fisher

4. Causation, Perception, and Dispositions: Towards *A Materialist Theory of the Mind* 52
Sharon Ford

5. Materialism, Then and Now 76
Peter Godfrey-Smith

6. Armstrong and Perception 92
David Rosenthal

7. Does *A Materialist Theory of the Mind* Give Us a Reply to the Knowledge Argument? 112
Frank Jackson

8. Closing (or at Least Narrowing) the Explanatory Gap 125
Katalin Farkas

9. Introspection and Distinctness: Armstrong and Shoemaker on Introspection 143
Ryan Cox

10. Armstrong's Just-so Story about Consciousness 176
Daniel Stoljar

11. Armstrong on Conation 195
William G. Lycan

12. Causal Conditionalism 211
David Braddon-Mitchell

13. A Materialist Reconception of the Mind 227
Amie L. Thomasson

14. Armstrong's Revenge 244
John Heil

Index 259

Acknowledgements

This volume arises from a conference held at the University of Sydney in August 2018 on '*A Materialist Theory of the Mind*: 50 Years On'. The occasion was both a celebration of David Armstrong's fêted book and an opportunity for critique and reflection on future directions for materialist theories of the mind. Ten of the fourteen chapters of this volume derive from papers delivered at the conference. The chapter by David Rosenthal was slated for the conference, but, in the event, Rosenthal was unable to attend. Likewise, Katalin Farkas had to withdraw from the conference, but has generously contributed a paper to the collection.

Of the other additional chapters, one is a paper by Armstrong himself, his last on the philosophy of mind, which is published here in English for the first time. It was first published in Russian in 2005 as 'Lewis i teoriy'a tozhdestva' [Lewis and the Identity Theory], *Vestnik MGU, Seriy'a 7: Filosofiy'a* [Moscow University Bulletin, series 7. Philosophy], issue 5: 79–84, translated by Vadim Vasilyev. The other addition, that of Peter Anstey and David Braddon-Mitchell, was written especially for this volume.

We would like to thank the Faculty of Arts and Social Sciences at the University of Sydney for financial support in the form of a Faculty Conference Grant in 2017 and the School of Philosophical and Historical Inquiry for a Seed Funding Grant. Without the generous financial support of the School and Faculty, the conference and this volume that issues from it would not have been possible.

Abbreviations

IKI	D. M. Armstrong, ‘Is Introspective Knowledge Incorrigible?’, *The Philosophical Review*, 72(4), 1963: 417–32.
MTM	D. M. Armstrong, *A Materialist Theory of the Mind*, London: Routledge and Kegan Paul, 1968. 2nd edition, London: Routledge, 1993.
T&M	Department of Traditional and Modern Philosophy, University of Sydney, 1973–2000.

Contributors

Peter R. Anstey is Professor of Philosophy in the School of Philosophical and Historical Inquiry at the University of Sydney.

David Braddon-Mitchell is Professor of Philosophy in the School of Philosophical and Historical Inquiry at the University of Sydney.

Ryan Cox is an Honorary Associate in Philosophy in the School of Philosophical and Historical Inquiry at the University of Sydney.

Katalin Farkas is Professor of Philosophy at the Central European University in Vienna.

A. R. J. Fisher is a Lecturer in Philosophy at the University of Washington.

Sharon Ford researches and studies at Victoria University, Melbourne.

Peter Godfrey-Smith is Professor of History and Philosophy of Science at the University of Sydney.

John Heil is Professor of Philosophy at Washington University in St Louis and Durham University and an Honorary Research Associate at Monash University.

Frank Jackson is Emeritus Professor at the Australian National University.

William G. Lycan is William Rand Kenan, Jr. Professor Emeritus at the University of North Carolina, and Distinguished Visiting Professor at the University of Connecticut.

David Rosenthal is Professor of Philosophy and Coordinator of Cognitive Science at the Graduate Center, City University of New York.

Daniel Stoljar is Professor of Philosophy in the School of Philosophy, Research School of Social Sciences, the Australian National University.

Amie L. Thomasson is Daniel P. Stone Professor of Intellectual and Moral Philosophy at Dartmouth College.

Introduction

Peter R. Anstey and David Braddon-Mitchell

In August 2018, a group of philosophers converged on the University of Sydney to discuss the philosophy of mind. The occasion was the fiftieth anniversary of the publication of David Armstrong's *A Materialist Theory of the Mind*. This volume contains, more or less, the proceedings of that conference, though with four conspicuous additions. The first is the chapter by David Rosenthal that was slated to be read at the conference, but in the event Rosenthal was unable to attend in person. The second is that by Katalin Farkas who accepted our invitation to the conference but who, in the end, had to withdraw. The third additional chapter is by the editors and convenors of the conference, Peter Anstey and David Braddon-Mitchell, and has been prepared especially to set out the historical context of this volume. The fourth, and by far away the most important, is a short paper by Armstrong himself entitled 'Lewis and the Identity theory'. This was first published in Russian in 2005 and is reproduced here in the original English for the first time.

It goes without saying that *A Materialist Theory of the Mind* (hereafter *MTM*) has been an enormously influential book: certainly, the most important book on the mind since Gilbert Ryle's *The Concept of Mind* appeared in 1949, and the most substantial contribution to Australian Materialism of the 1960s. It therefore seemed appropriate in 2018 when it turned fifty to reflect on its origins, its influence, and its enduring legacy.

The first four chapters of this volume are largely historical in orientation. We, the editors, open the collection with a chapter that aims to set out the proximal and distal issues, both philosophical and polemical, that gave rise to this, Armstrong's most important book. It addresses the specific themes of Armstrong's materialism: the centrality of causation, his attitudes toward the Identity theory and conceptual analysis, as well as his response to Behaviourism. The chapter then turns to the broader theme of where *MTM* fits in the history of twentieth-century materialism and the various methodological manoeuvres that materialists pursued.

Chapter 2, Armstrong's 'Lewis and the Identity Theory', recounts the origins of Lewis's seminal paper 'An Argument for the Identity Theory' which appeared in 1966, two years before *MTM*. It also situates Lewis's theory in relation to

Peter R. Anstey and David Braddon-Mitchell, *Introduction* In: *Armstrong's Materialist Theory of Mind.* Edited by: Peter R. Anstey and David Braddon-Mitchell, Oxford University Press. © Peter R. Anstey and David Braddon-Mitchell 2022.
DOI:10.1093/oso/9780192843722.003.0001

Armstrong's own theory and in so doing adds an important seam to the familiar narrative of how the Causal theory of mind emerged in the years immediately after the publication of *MTM*. The historiography, created and propagated by Armstrong himself, has always rightly emphasized the priority of U. T. Place and J. J. C. Smart in the emergence of *MTM*, but until now we have not had such a detailed account from Armstrong himself of Lewis's 'causal role' in the first wave of thinking about the mind post-*MTM*.

Chapter 3, Anthony Fisher's 'The Two Davids and Australian Materialism' continues the theme of the relation between Armstrong and Lewis, only Fisher casts the net far wider. He begins by arguing that there were at least two different lines of influence from early twentieth-century Behaviourism to the Identity theory: one through Logical Positivism (Feigl) and the other through ordinary language philosophy, the latter involving Place and Smart, and Lewis and Armstrong. It was Armstrong and Lewis who were to have a profound influence on subsequent developments in analytic philosophy, both methodologically and in metaphysics, and it is for this reason that Fisher devotes the rest of his chapter to tracing the origins of their theories of mind and the manner in which the emphases and methodologies deployed in developing their respective theories became the hallmarks of their wider impact on analytic philosophy. Fisher argues that in Armstrong's case, the philosophical methodology employed in *MTM* presages the argument forms that he deploys in his later work in analytic metaphysics. Fisher furnishes a similar case for Lewis's later work and its influence, claiming that he arrived at his doctrine of theoretical terms by working out his philosophy of mind and then applied this doctrine to a wide range of philosophical topics from metaphysics to value theory.

Chapter 4, Sharon Ford's 'Causation, Perception, and Dispositions: Towards *A Materialist Theory of the Mind*' also explores the origins of *MTM*, though this time through a clutch of fascinating and hitherto unexamined materials, namely, Armstrong's undergraduate essays and early unpublished papers. What Ford establishes is the striking degree of continuity between Armstrong's earliest philosophical writings and the first works of his philosophical maturity. The themes of perception, sensation, causation, dispositions, and properties were taking shape from his undergraduate days under the tutelage of John Anderson, and continued to develop at Oxford, Birkbeck, and Melbourne, before the decisive move to Sydney University. And a number of distinctive Armstrongian doctrines were adumbrated during these early years.

Clearly the nature of perception had a central role in both the pre-history and in the articulation of Armstrong's materialist theory of the mind in *MTM*, however, as Peter Godfrey-Smith argues in Chapter 5, there is a sense in which perception is peripheral in contrast to the core mental states that constitute the belief-desire nexus. Moreover, when it comes to beliefs, it seems far too simplistic to claim that, on its own, any belief is apt to bring about a determinate behaviour; the situation is far more complex. However, Godfrey-Smith argues, an overlooked theme in the work of Frank Ramsey can ameliorate the situation, namely, 'belief-desire profiles', including degrees of belief. These provide an explanatorily richer way of developing a causal theory of the mind than atomic beliefs, or Armstrong's later conception of beliefs conceived as maps. Then, in a short final section, Godfrey-Smith turns to arguments for and against materialism, challenging Armstrong's appeals to the principle of parsimony, and offering as an alternative to materialism a form of neutral monism which posits that there is a non-physical, non-mental basic reality.

Perception is also the focus of David Rosenthal's Chapter 6, 'Armstrong and Perception'. Rosenthal challenges Armstrong's view that sensations, sense impressions, and mental images lack any qualitative nature. He argues that all of Armstrong's arguments against qualitative mental perceptions presuppose a criterion of consciousness, namely, exclusive first-person epistemic access on the part of the percipient. However, if mental qualities are thought of in terms of a fine-grained 'quality space', and we focus on Armstrong's conception of perception as the formation of a belief, ignoring the consciousness criterion, then there is, according to Rosenthal, a coherent account of mental qualities available that is consistent with materialism.

Frank Jackson's Chapter 7 leverages Armstrong-influenced discussions of colour and of perception to provide a reply to his (Jackson's) famous version of the knowledge argument against physicalism. Jackson argues that Armstrong's treatment of perception in terms of the acquisition of belief is along the right lines but needs to be modified. We should think of perceptual experiences as putative responses to the world that by their very nature urge upon us a view about how the world is. Experiences as of an object's surface being a certain colour are, accordingly, states that urge that the object's surface is a certain way, and an object will have that colour just if its surface is the way in question. But what is that way? Jackson argues that the way in question is having a surface property that stands in a certain location in a complex similarity space. He arrives at this view by explaining why he dissents from Armstrong's well-known version of physical realism about colour. This view of Armstong's holds that colours are (in the case of coloured surfaces) either intrinsic physical properties of the surfaces, or are dispositional properties closely associated with the surfaces, such as reflectance profiles. However, Jackson's view is compatible with a broadly physicalist outlook in the following sense: the properties that stand in the relevant locations in the similarity space will be physical properties, reflectance profiles as it might be. The implications for Mary in the black and white room are, first, that she can know about colour. This is because she can know that the surfaces of the things we are in contact with stand in certain patterns of similarity—perhaps the patterns of similarities expressed in a colour cube—by virtue of her vast knowledge of physical properties. If the relevant physical properties are indeed reflectance profiles, this knowledge will come from her knowledge of how the reflectance profiles of the objects we see are alike and unalike. The second implication is that she knows about colour experiences in the following sense: she knows that those outside the black and white room are often in states that by their very nature urge on them that the objects they see are alike and unalike in the ways definitive of being one or another colour. The information she can only get by laborious calculation, they get immediately by simply looking. The upshot is that Mary learns nothing new when she leaves the black and white room. What happens is that she is altered causally so as to be in states whose intrinsic nature directly (no need for calculation) urge her to take things (perhaps a rose in front of her, and its leaves) to occupy such and such a location in a similarity matrix. This is a remarkable thing to happen to her, and gives her an ability to manipulate, examine and recognize the states that play these very distinctive informational roles that she previously lacked. It provides her with a kind of mental laboratory whose richness makes the Lewis-Nemirow approach far more plausible than before. We have, accordingly, a version of the Lewis-Nemirow reply to the Knowledge Argument—what Mary gains when she leaves the black and white room is know-how, rather than propositional knowledge. It is, though, a version which arguably deals with what worries so many about that reply, the sense that it does not explain the very distinctive nature of what happens to Mary when she leaves that room.

In her Chapter 8 on the causal theory of the mind and the problem known as the explanatory gap (following Joseph Levine), Farkas revisits the problem of the explanatory gap that is supposed to open up when considering identity statements between physical and mental phenomena. She shows that the question asked in the original formulation of the explanatory gap was this: 'Why is this phenomenal character, rather than any other, attached to this physiological process?' And she goes on to argue that this question can be answered, because there is a natural fit between the phenomenal character of experiences and their functional roles. For example, pains feel inherently unpleasant, and this explains why they cause avoidance behaviour, and moreover, the felt location and intensity of pain indicate the location and extent of damage to the body. After discussing some other possible gaps in our understanding of the relationship between the mental and the physical, Farkas concludes that the fit between functional role and phenomenal character goes a long way, though probably not the whole way, to closing the explanatory gap.

Ryan Cox's Chapter 9, 'Introspection and Distinctness: Armstrong and Shoemaker on Introspection' deals with Armstrong's self-scanning theory of self-knowledge. Armstrong held the view that when we know that we are in some mental state, it's a question of scanning our own internal states. Cox claims that this was thought by Armstrong to follow from a combination of a traditional view of perception, and something like the central-state materialist view that mental states are internal states of the brain. This made Armstrong very keen to reject the other traditional view that we have incorrigible access to our own mental states. He rejected this, according to Cox, because he thought if our access were incorrigible, this would be inconsistent with this combination of views. This leads to Armstrong's own argument for why we should expect our beliefs about our own mental states to be fallible, which hangs on the idea that our beliefs about mental states, and the states they are beliefs about, are distinct, together with Hume's dictum that there are no necessary connections between distinct existences (which there would have to be, it might seem, for this class of beliefs to be incorrigible). Cox's chapter considers the various ways in which we might cash out the idea of distinctness, and contrasts Armstrong's view with Shoemaker's view, according to which they are not distinct.

The theme of Armstrong on introspection continues in Daniel Stoljar's Chapter 10, 'Armstrong's Just-so Story on Consciousness'. In this chapter, Stoljar argues that there is a fascinating tension between Armstrong's purposed self-scanning model of introspection, and his account of the function of consciousness. The account of the function of consciousness is roughly that consciousness is necessary for purposive thought: in order to think things through, we must keep track of and be aware of our various thoughts. Thus, he claims it's necessary that, if we have purposive thought we are conscious (as Stoljar points out, this doesn't make it a biological function, but rather a precondition for a function, but that needn't detain us here). But then it looks like it's a kind of constraint on rationality that we are aware of our thoughts—that's what it takes to have rational thought. The price, for Armstrong, of thinking this is that the idea that it is a constraint on rationality that we are in touch with our thoughts is precisely a leading contender for capturing the *denial* of a self-scanning view. Stoljar identifies just such a problem in later discussions of introspection, and concludes with a general worry for any attempt to give an account of consciousness in terms of being informationally sensitive to our own states—surely such sensitivity could be achieved at the sub-personal level.

In Chapter 11, William (Bill) Lycan turns his attention to Armstrong on conation. The principal focus of his chapter is the so-called content specification or grain problem for the content of desires. Anyone who has read stories about the cunning way genies in bottles make life bad for their masters in granting wishes will be familiar with the problem. The canonical expression of a desire almost without exception leaves open that the proposition expressed could be made true without your getting what you want. You want a new shirt; you get one that you are allergic to. Or one that is so hideous you are reviled by your friends and colleagues. Or that is so expensive you can't eat. Or any other of a potential infinity of defeaters. How to specify the desire more precisely? It's not that there is, necessarily, a particular shirt under particular circumstances that you want. You want a shirt, you don't care how, just so long as, well, you thereby get what you want.

Lycan explores the strategy of specifying the content of a desire as a kind of conditional. Your desire has a kind of conditional content, with the antecedent of that conditional involving the conditions under which the consequent is desirable. He compares his own preferred version of this strategy with others, and finds resources in Armstrong's account of conative states to solve the problem. As an afterword, he also finds in Armstrong's work a solution to the so-called 'disjunction problem' for co-variational accounts of intentional content: a solution available long before the problem gets its most influential formulations. The problem is that if the content of mental states is to be understood as their co-varying with worldly states, representational states of the mind co-vary with much more than the intuitively intended states. If I have a state that represents cats, it co-varies not only with cats but with cat like appearances, and misleading evidence of cats, as much as it does with cats. And yet intuitively it is a state that represents cats, not possums on a dark night. The strategy Lycan finds in Armstrong goes via the idea of the purpose of the state: not the teleofunctions, as one modern account would have it, but the *agent's* purposes for either of cats or cat representations.

The next chapter in the collection, Chapter 12 by David Braddon-Mitchell, returns to the central thesis of *MTM*: the truth of materialism. In this chapter, Braddon-Mitchell develops a new response to the Conceivability Argument against physicalism. It is a response designed to have the benefits of an earlier response—the so-called 'conditional concept' strategy. But that response moves away from the pure causal theory (or analytic functionalism) of Armstrong to a view where the causal theory, while perhaps true, is no longer an analysis. The new view uses some thoughts of Armstrong about causation to develop a conditionalism about causation which makes it possible for the casual theorist to accept that they can coherently conceive of a physical duplicate of the world which is a 'zombie' world—i.e., contains no experience—while still holding that such a world is metaphysically impossible. The core idea is that, for all we know, a priori causation is a non-physical matter. The fact that many philosophers think it is a non-physical matter supports this. Thus, while analytic functionalists think they know a priori that a casual duplicate of actuality is a mental duplicate, they do not know a priori that a physical duplicate is a causal duplicate. How then can they hold that a zombie world is nevertheless impossible? Because they might hold that it's a priori that if causation is actually a purely physical matter, then necessarily every physical duplicate is a causal duplicate, and also that causation is in fact actually a purely physical matter. But they cannot know the last thing—the antecedent of the conditional—a priori, since the physical status of causation, while possibly true, is not a priori true.

Amie Thomasson's Chapter 13 continues the theme of the analytic or a priori status of the causal theory. She suggests that rather than conceiving of this analytic status as something that is *discovered* by conceptual analysis, we should think of it as an exercise in what has become known as conceptual ethics. Just as some behaviourists took their view not to expose the concept we already had, but as a suggestion for conceptual reform, we might take Armstrong's causal theory as a suggestion for conceptual reform in the light of scientific and other discoveries which might make us doubt whether there is anything in the world corresponding to some of our pre-reform concepts of the mental. She goes on to argue that while Armstrong certainly never explicitly understood what he was up to in these ways, it's certainly clear that in his later writings he became a little unhappy with the idea that the causal theory was a straightforward analytic truth about existing concepts—even though he had no clear idea about what the alternative might be.

In the final chapter of the volume, John Heil distinguishes two kinds of functionalism: what he calls, after Block, 'functional specifier functionalism' (often now called 'realizer function'). The first of these views was Armstrong's and Lewis's view, and remains the dominant strain in Australian philosophy of mind. But the latter view went on to become dominant in the US. Heil diagnoses a range of problems for philosophy of mind in the last forty years as a consequence of that latter view's dominance. In particular, issues about mental causation go away on the former view, as mental states literally are the causally efficacious states that play roles, whereas, on the latter view, mental states are abstract such as 'the state of having a state that plays a role' whose causal efficacy is dubitable. Heil argues that problems for the former view can be dealt with by modifying it to incorporate his own account of dispositions, according to which dispositions are identical with their categorical bases. The result is a view that is more thoroughgoingly first order. Mental states are first-order states rather than abstract second-order states like the "the state of being in some state that has some specifications". In addition, dispositions are equally first order. He concludes by arguing that this modification has the potential to support a straightforward identification of experiential states with neurological states that avoids the pitfalls of previous attempts to do this.

Some Notes on Terminology

Materialist theories go by many names, and in some cases there is much confusion. Armstrong initially called himself a central state materialist. Lewis, Braddon-Mitchell, and Jackson, and many others have called themselves analytic functionalists or common-sense functionalists. Armstrong later called his theory the Causal Theory of the Mind. To a first approximation, these are very similar theories. As for Smart, he did not intend his view to be a type identification of mental states with particular types of central states (except contingently). He thought that we learn that our mental states are, in us, a kind of central state because of the theoretical roles they play (though the emphasis in causation is not as explicit as in Armstrong or in avowed analytic functionalists). Explicit analytic functionalism is surely the same view: explanatory causal theories in which the mental state terms are defined via a Ramsey sentence in terms of what they are apt to cause (and be caused by). Of course there are differences in emphasis, but no one has gone into the details of so many mental states as Armstrong.

But, of course, 'Functionalism' is a word with many meanings, and if we ask, 'Is Armstrong a functionalist?', we had better be clear. If a realizer analytic functionalist, yes. But many doctrines called Functionalism are not Armstrong's view. Even role analytic functionalism is surely not Armstrong's view. Role analytic functionalism is the view that mental states are second-order states: the states of having certain states play causal roles. So mental states do not themselves actually cause things, they only give information that certain causings are taking place. Armstrong thought that the mental was powerful: the mental states actually do causing. Still less is Armstrong a functionalist if what we mean is the doctrine that Putnam sometimes held: that mental states are a kind of abstract state brought into existence by material or other things configured in a certain way. For Putnam, multiple realizability means that the very same kind of state can exist brought into being or otherwise supervening on many things. For the realizer functionalist, multiple realizability means that many different states deserve to be called mental because of how they cause; there's no extra mental abstract reality. That Armstrong is in this camp should be clear from his discussion of poison. Many different physical substances are poisons. They are physically different, but they deserve to be called poisons because, in the right context, they kill. But it is not that there is some extra abstract feature which is the poison, which is present each time one of these different substances is present.

1

A Materialist Theory of the Mind in Context

Peter R. Anstey and David Braddon-Mitchell

The publication of David Armstrong's *A Materialist Theory of the Mind* (hereafter *MTM*) was the high-water mark of a new movement in the philosophy of mind that began in the late 1950s in the work of U. T. Place and J. J. C. Smart at the University of Adelaide and continued with Brian Medlin and Armstrong himself. Yet it was more than this: it was also the first sustained application of a new philosophical methodology that was at once deferential to the empirical sciences while at the same time re-establishing a central role for metaphysics. Thus, *MTM* was not a stand-alone work; rather it was the longest and most thorough expression of a philosophical trajectory that had been gaining momentum within analytic philosophy throughout the 1960s.

Nor was this an exclusively Australian affair: a parallel development was underway in the United States in the work of a young David Lewis, though Armstrong was, apparently, unaware of Lewis's work at the time of writing *MTM*. Moreover, in *MTM* we find the coalescence of a series of philosophical issues that had a life of their own: these issues include materialism, the nature of causation, the nature of dispositions, perception, secondary qualities, and consciousness. Armstrong's project implicated each of these issues in the development of a materialist theory of the mind, yet he and others within his ambit had already addressed, and would continue to work on, each of these issues independently of this unified project in the philosophy of mind.

The history of this moment in twentieth-century philosophy has been told a number of times and from different perspectives. What we aim to do here is to tell it again, in two complementary parts, though now with the benefit of fifty years of hindsight. First, we will discuss the more proximal influences on Armstrong's book, then we will prescind to a more abstract level and talk about how the work fits into the narrative of materialist and non-materialist philosophy of mind in the twentieth century.

All of the major players in this moment have now passed away, yet their influence lives on, and the occasion of the fiftieth anniversary of the publication of *MTM* has provided an opportunity for fresh reflection on this exciting period of our not-too-distant philosophical past.

Peter R. Anstey and David Braddon-Mitchell, A Materialist Theory of the Mind *in Context* In: *Armstrong's Materialist Theory of Mind.* Edited by: Peter R. Anstey and David Braddon-Mitchell, Oxford University Press. © Peter R. Anstey and David Braddon-Mitchell 2022. DOI:10.1093/oso/9780192843722.003.0002

1. Proximal Influences

The first publication in which Armstrong discussed his materialist theory of the mind appeared in 1966 and was entitled 'The nature of mind' (Armstrong 1966). To be sure, much preliminary work had already been undertaken in his direct realist account of perception in *Perception and the Physical World* (Armstrong 1961) and *Bodily Sensations* (Armstrong 1962) which, in turn, built on his work on *Berkeley's Theory of Vision* (Armstrong 1960). Thus, by the time he turned to the mind, he already had a fully-fledged theory of perception which included a new analysis of bodily sensations. Indeed, a careful survey of Armstrong's early writings, even from his undergraduate days, shows that a path towards a materialist theory of the mind was taking shape (see Ford, Chapter 4 in this volume). Nevertheless, it was in 'The nature of mind' that the project was first adumbrated and it is important that we unpack the dialectical structure of this short article in order to see what issues were driving Armstrong as he began to compose *MTM*.

1.1 Materialism

'The nature of mind' opens with an appeal to the epistemic authority of science and a statement of the general premise on which the project was to be founded: 'we can give a complete account of man in *purely physico-chemical terms*' (Armstrong 1966, 1). This premise was undergirded by the authority of the current state of science. Armstrong admitted that future developments in science may present problems for his foundational premise (Armstrong 1966, 1), but given the current state of play, it was a secure starting point for 'what may be called a Materialist or Physicalist account of the mind' (Armstrong 1966, 2). We note here that already Armstrong is using 'physicalist' as a synonym for 'materialist'.

These points were soon incorporated into *MTM*. For example, the focus on the authority of science as the foundation for the theory is repeated on the opening page of *MTM*:

> Increasing scientific knowledge has largely answered the first objection [How could life be purely physical?]. It is now very probable, even if not certain, that life is a purely physico-chemical phenomenon. We do not need to postulate 'vegetative souls' or 'vital entelechies' to explain life. (*MTM*, 1)

Furthermore, the first four chapters of *MTM* are concerned with arguing for the superiority of a materialist theory of the mind over various non-materialist alternatives, especially dualism about the mind and body. And the fourth chapter of *MTM* provides a general argument against any non-materialist theory of the mind from the success of the physical sciences in all areas apart from psychology. The argument developed here, 'the Argument from the supremacy of physics' (*MTM*, 50), is really an argument from inference to the best explanation, and Armstrong supplements it with considerations from G. E. Moore and the notion of a general consensus (*MTM*, 51–2).

It is important to note, however, that Armstrong's early emphasis on the authority of science as a bulwark for the credibility of a materialist theory of the mind drops away relatively quickly. His under-appreciated 'The headless woman illusion and the defence of materialism' (Armstrong 1968b), which was written as a kind of postscript to *MTM*,[1] is an attempt to explain the fact that mental images are not experienced as processes occurring in the brain. Here the aim is to block the objection that the phenomenology of the mental counts against materialism about the mind. Moreover, the second part of his 'Epistemological foundations for a materialist theory of the mind', written by late 1971,[2] provides an account of 'rational consensus' which he uses to argue for a kind of 'epistemic credit rating' that ranks common sense and science highly. Here the key precursors are G. E. Moore and J. J. C. Smart respectively: 'It seems to me that Moore and Smart, taken together, indicate to us the proper epistemological foundations for philosophy' (Armstrong 1973, 185). And he goes on to defend the program for developing a materialist theory of the mind in the light of objections, such as that of the phenomenology of introspection, appealing again to the headless woman illusion (Armstrong 1973, 190–1). But that is really the last we hear of it. From the early 1970s, the need to defend the materialist nature of his theory of the mind seems to fade away and other issues come to the fore.[3] The most important of these is causation.

1.2 A Causal Theory

That the notion of causation was central to Armstrong's materialist theory of the mind from the outset is evident in the way he sets out his response to Behaviourism in 'The nature of mind'. The suggestion is, 'Perhaps mind can be defined not as behaviour, but rather as the inner *cause* of certain behaviour' (Armstrong 1966, 7). Yet in this early paper the language is somewhat tentative. By the publication of *MTM*, the causal analysis of behavioural dispositions had become a hallmark of the theory and yet causation was still far from being the defining characteristic of the theory.[4] All of this was to change when Armstrong read David Lewis's paper on the identity theory.

---

[1] See 'Self-Profile', Armstrong (1984, 31).

[2] Armstrong to David Lewis, 25 November 1971, David Lewis Papers, C1520, 'Armstrong, D. M.', Box B-000659 Folder 15, Princeton University Library.

[3] This is not to say that Armstrong ceased to defend materialism as an ontological thesis. See, for example, 'Naturalism, Materialism and First Philosophy', 1975, reprinted in Armstrong (1980, 149–65).

There is no mention of David Lewis's study, 'An Argument for the Identity Theory' (Lewis 1966) in *MTM*, but Armstrong had grasped its importance for his own theory by around 1971.[5] He first mentions it in his 'Epistemological foundations' paper and this presages an important change of emphasis in the manner in which Armstrong framed his whole project in the philosophy of mind: 'David Lewis has suggested a more general, and even more helpful, formula for capturing the concept of a mental process or state. He proposes that all mental concepts are concepts of that which plays a certain causal role in the physical behaviour of the organism' (Armstrong 1973, 180).

Lewis's notion of a causal role was a key conceptual addition and Armstrong was quick to appropriate it; and this allowed him, in turn, to view both the input and output aspects of causal roles to have the same standing. Indeed, it is from this point that he begins to call his theory the casual theory of the mind: '[w]e may call Lewis' and my account of the mental concepts the "Causal" account of mental concepts' (Armstrong 1973, 181).

What is interesting is the speed with which the shift from the materialist theory to the causal theory took place. The broader polemical origins of the theory, which had seemed so important to Armstrong in the 1960s gave way to the core explanatory power of the theory—the notions of causation and causal roles—and the nomenclature followed suit. Thus, the next important instalment in Armstrong's oeuvre is a paper simply entitled 'The causal theory of the mind'. This widely cited and reprinted paper drops the emphasis on materialism almost entirely and brings the notion of causation to the fore. It is this that now provides the counter to those with an inclination to dualism about the mind (Armstrong 1977a; 1980, 24–5). Reflecting on this in his Preface to the second edition of *MTM* in 1992, Armstrong says, 'Later on I just called it the Causal Theory (see my paper "The Causal Theory of the Mind", 1977). Nowadays I sometimes speak of it as the Causal version of Functionalism' (Armstrong 1993, xiv). Then, in his historical treatment of the philosophy of mind, *The Mind-Body Problem: An Opinionated Introduction*, which appeared in 1999, virtually all traces of the materialism have disappeared. The theory is set out as 'The Causal Theory—Armstrong and Lewis' and as a natural development of The Identity Theory, which is the title of the previous chapter.

---

[4] Armstrong later claimed that C. B. Martin was an important influence on his stress on causation. See Armstrong (1989, 11–12).

[5] Armstrong apparently had not read Lewis's paper when he wrote 'Dispositions are causes' in the wake of the publication of *MTM* in 1969. Neither his interlocutor Squires (1968) nor Armstrong's reply (Armstrong 1969) mention Lewis and the notion of a causal role.

1.3 The Identity Theory

Armstrong seems never to have wanted to be identified with the identity theory of mind. The label had been in use since early in the twentieth century and was very much on the scene in Anglo-American philosophy by the mid-1950s. It is used, for example, in Herbert Feigl's 'The Mind-Body Problem in the Development of Logical Empiricism' of 1950. Armstrong, however, does not align himself with the theory in 'The nature of mind' in 1966 and in *MTM* he calls his own view central-state materialism, a term adapted from Feigl's earlier 'Central-state theory' (*MTM*, 10; Armstrong 1999, 67; see Feigl 1958) and one that connotes the important role of the central nervous system to the theory. Chapter 6 of *MTM* is entitled 'The Central-State Theory' and this chapter sets out the central doctrine of the book.

In his reflections on the identity theory in the Preface to the second edition of *MTM*, Armstrong speaks quite candidly about his view of the identity theory:

> Perhaps because of the way that my argument took off from the positions of Place and Smart, positions that were very close to each other, I am usually represented in the literature as holding an Identity theory. I was never too keen on the phrase and did not use it to describe my view. (Armstrong 1993, xiv)

However, he does concede that when viewed from the functionalist's application of the token/type distinction, the position set out in *MTM* was a version of the identity theory: 'if the Identity theory is *defined* as a type–type identification of the mental with the physical, then I did hold an Identity theory' (Armstrong 1993, xv).

It was not long, however, before Armstrong also dropped the name 'central-state materialism' and sat more comfortably with the appellation 'causal theory of the mind'. As we have mentioned above, in his opinionated introduction to the mind-body problem published in 1999, he actually juxtaposes the Armstrong-Lewis causal theory with the identity theory of Place and Smart, implying that the causal theory was actually a subsequent developmental stage beyond the identity theory.

One very important factor in this shift was Armstrong's quick acceptance of the critique of type-type identities rather than token-token. Again he is very explicit about this: 'when this criticism . . . was put to me I saw the point and accepted it immediately' (Armstrong 1993, xv). In fact, in his 'Self-Profile', Armstrong claimed that this was 'my greatest omission' and that '[a] few moments' thought about the matter would have shown me that there were problems in identifying, say, the type pain with a certain physiological type' (Armstrong 1984, 32). He intimates there that one of the factors that blinded him was his enthusiasm for Brian Medlin's analogy of the gene-DNA model for the mental state-brain state relation, though it must be said that his attraction to that analogy never waned and he still endorsed it in 2005 in 'Lewis and the identity theory' (Armstrong, Chapter 2 in this volume; Medlin 1969). It seems most likely that Armstrong gave the issue 'a few moments' thought' when he read a draft of Lewis's article 'Mad pain and Martian pain' (Lewis 1980) in late March/early April 1978 (a full decade after the publication of *MTM*), for, in a letter to Lewis of 13 April of that year he wrote:

> It is not clear to me that what I argued for in M.T.M. was a type-type theory. (I was simply numb and vague about the whole matter.) Certainly, I wanted something stronger than a token-token theory—because that seems to allow Dual-Attribute theories. But some interpretations at least, of the phrase 'type-type theory' seem to be too strong. Suppose that pains which were phenomenologically indistinguishable, and whose causal roles were just the same, were correctly identified with a disjunction of two very different sorts of brain-process. (I think this is conceivable.) Would this be a 'type-type' identity? If so, O.K.—I would not object to being called a 'type-type' theorist. But I am dubious about any theory which forces us to say that each type of mental state is identical with just one type of brain-state. (As, I take it, you are too.) (Armstrong to Lewis, 13 April 1978)[6]

These remarks suggest that Armstrong perhaps had sympathies with what later became known as a domain-specific type-type identity theory: one which makes restricted type identity claims, while not thinking that there is a cleanly identifiable physical type to which pain (for example) in general is identifiable. Thus, pain-in-humans might be identified with one physical type, and pain-in-dolphins (or even pain of another kind in humans) with another (see Jackson, Pargetter, and Prior 1982). The impossibility of an absolutely general identification of any mental state with a physical state becomes clear in later versions of causal functionalism, because since a causal functionalism says that it is sufficient to be an instance of a mental state to play certain causal roles, if there is a possible state of affairs in which something non-physical were to play that causal role, that role player would count as an instance of the mental state. Thus, materialism about the actual world is the combination of the causal theory and the empirical claim that all actual causation is material: it does not follow directly from the analysis. If mental states in absolute generality (rather than just in humans, or even in the whole of actuality) were type identical with physical states, then the causal theory would have to be false. This is because the causal theory claims that in worlds where *causation* is non-physical, mental states are not instances of a physical type.

---

[6] David Lewis Papers, C1520, 'Armstrong, D. M.', Box B-000659 Folder 15, Princeton University Library.

1.4 Conceptual Analysis

From the outset, Armstrong attempted to justify the methodology of his philosophical project in the philosophy of mind. This required more than merely founding his basic premise on the authority of the sciences, it also required a statement and justification of the philosophical techniques and practices that he deployed. The most important among them was conceptual analysis which he openly traced back to the influence of Oxford philosophy and, in particular, Gilbert Ryle. Reflecting on his time at Oxford in his 'Self-Profile', published in 1984, he says: 'I suppose that what I eventually took from Oxford was the idea that there was such a thing as conceptual analysis, that it could be important and useful in philosophy—I could never accept that it was the whole of philosophy—and that linguistic investigation could in some degree trace out the structure of our concepts' (Armstrong 1984, 11).

Armstrong admits that a large part of *MTM* is conceptual analysis and he defends this in Chapter 4 against potential critics by pointing out that the exposition of the doctrine in *MTM* is not based 'solely on conceptual analysis' (*MTM*, 51). He claims that the development of the theory involves two stages: (1) 'a certain logical analysis of mental concepts'; and (2) an empirical investigation that identifies mental states with 'physico-chemical states of the brain' (*MTM*, 91). It is the first stage that is the concern of the book and he views Part Two, that is 'the intellectual centre of this book', and bears the Rylean title 'The concept of mind', as 'a detailed working out of the Causal analysis of the mental concepts' (*MTM*, 91). He reiterates this in the article 'Epistemological foundations' where he says—though now in the context of Lewis's notion of a causal role—that *MTM* provides 'analyses of the individual mental concepts' (Armstrong 1973, 181).

A more focused defence of his philosophical methodology in *MTM* and in the philosophy of mind more generally, appears in 'The Causal Theory of the Mind' of 1977. Armstrong first gives a thumbnail sketch of the rise, fall, and moderate reinstatement of conceptual analysis in the methodology of Western philosophy. He claims that one weakness of those who ostensibly practised pure conceptual analysis was the surreptitious importing of substantive metaphysical doctrines to guide the process. Ryle's commitment to behaviourism in *The Concept of Mind* is the salient example (Armstrong 1980, 17). He goes on to contend that the philosopher brings to a problem like the mind-body problem two specific skills: (1) the evaluation of arguments; and (2) the analysis of concepts. The moderate reinstatement of conceptual analysis amounts to this: the analysis of concepts is to be done in the light of explicit metaphysical commitments—commitments to first-order questions—and in deference to the physical sciences (Armstrong 1980, 18). All of this is by way of an introduction to the next section of the paper which is entitled 'The concept of a mental state'.

Reflecting on all of this in his Preface to the second edition of *MTM*, Armstrong claimed:

> I put forward the Causal theory as a theory about concepts, as a set of logical analyses. This was probably the result of the Oxford part of my time as a philosophy student (1952–4). I would be much less dogmatic now. Perhaps it is just a theory about the mental instead of an analysis of our concepts of the mental. (Armstrong 1993, xv)

Interestingly, in his penultimate writing on the philosophy of mind,[7] which is published in this volume in English for the first time, Armstrong makes a similar claim about the influence of Oxford philosophy on David Lewis: 'One thing that David took from Oxford was an enduring respect for conceptual analysis . . . one concept that certainly caught his attention in Oxford was the concept of mind' (Armstrong, Chapter 2 in this volume, page 24). In his assessment, while his own view of the efficacy of conceptual analysis of the mental had weakened, Lewis remained committed to the end: 'though I am less sure than I once was that the account is no more than a conceptual analysis. (David, though, never wavered in his faith in conceptual analysis here.)' (Armstrong 1993, xv).

1.5 Behaviourism, Ryle, and Dispositions

Armstrong's example of Ryle's covert commitment to behaviourism was not merely a well-chosen example of first-order questions informing conceptual analysis, for it provided one of the primary terms of reference for the framing and development of his whole project in the philosophy of mind. In fact, so central was it that in his earlier 'The nature of mind' he went so far as to frame the project in Hegelian terms. The thesis is the classical conception of the mind as an 'inner arena'. The antithesis is behaviourism (as found in Ryle). And the synthesis is a conception of the mind as inner states apt to bring about certain types of behaviour (Armstrong 1966, 10). Interestingly, Armstrong reproduces this as the opening paragraph of Part Two ('The Concept of Mind') of *MTM* (*MTM*, 129).

Here is the dialectic as set out in 'The nature of mind'. Behaviourism posits that the mind is merely behaviour and that there is no inner arena, but there is the obvious objection that we seem to experience inner mental processes without any outward display of behaviour. Ryle answered this challenge by appealing to dispositions to behave. Rylean dispositions, however, amounted to nothing more than being liable to undergo a particular change, such as manifesting a certain behaviour. Yet this can be taken a step further; we can identify dispositions with mental states that are 'logically tied to behaviour' (Armstrong 1966, 7). On this view, mental states are the actual underlying causes of behaviour. These, in keeping with the physicalist premise, are to be identified with 'purely physical states of the central nervous system' (Armstrong 1966, 8). Here the brittleness of glass or the elasticity of rubber provide useful analogues, for the scientist can identify the molecular structures and forces that constitute the dispositions. So, the claim that the mind is merely behaviour is unsatisfactory, and we therefore need dispositions. But dispositions as mere propensities to undergo change are not enough; we need a robust causal theory of definitions that is consistent with dispositional analyses as found in the empirical sciences. Hence, mental states are states of the brain that are apt to cause certain behaviour.

---

[7] The final chapter of his last book, *Sketch for a Systematic Metaphysics*, is entitled 'Mind'. See Armstrong (2010, 105–15).

Now, three interesting points are worth making in relation to the place of behaviourism, Ryle, and dispositions in the development of Armstrong's materialist theory of the mind. First, while Ryle himself disavowed behaviourism, it is clear from very early on that Armstrong read him as committed to the doctrine. Armstrong purchased *The Concept of Mind* in 1951 and his copy contains many annotations. One early marginal note is striking; alongside Ryle's statement, 'Overt intelligent performances are not clues to the workings of minds: they are those workings', he wrote 'Ryle's Behaviourism' (Armstrong's copy of Ryle 1949, 58).

Second, it is worth noting the parallels between *MTM* and Ryle's seminal *The Concept of Mind*.[8] It is not simply the role of Ryle's theory in the polemical structure of *MTM*: there are also some obvious parallels in the contents. We have already noted that the title of Ryle's book is the title of the central Part of *MTM*. In addition, the opening two chapters of Part Two on 'The Will' correspond to Ryle's Chapter III on 'The Will' and both books have material on emotions (Ryle 1949, Chapter IV; *MTM*, Chapter 8, §III) and introspection (Ryle 1949, Chapter VI; *MTM*, Chapter 15). It is not that *MTM* grew out of *The Concept of Mind*, but rather that one cannot fully appreciate the novelty and significance of *MTM* without getting one's bearings from the leading work on the philosophy of mind of the previous generation. No doubt it was with some glee, therefore, that in his final paper on the philosophy of mind, Armstrong quoted Lewis's comment: 'I am an ex-Rylean, and I retain some part of the Rylean legacy' (Armstrong, Chapter 2 in this volume, page 25).

Third, and finally, while the influence of Place and Smart has been widely acknowledged, it ought to be pointed out that early on Armstrong acknowledged the influence of C. B. (Charlie) Martin on his thinking about causation and dispositions.[9]

---

[8] Armstrong's debts to Ryle predate 'The nature of mind' and *MTM*. Thus, Ryle's (1951) paper 'Feelings' was 'the starting point' for Armstrong's *Bodily Sensations* (1962). See Armstrong (1962, vii). Anthony Fisher alerted us to this point.

2. The Place of Armstrong's View in the History of Materialism

So much for the proximal influences on Armstrong's materialism, where do his theory and other theories of a similar kind fit into the history of materialism over the last century, and how does it contrast with non-materialist theories?

There are many ways to understand the motivational impulse behind materialism, and these different motivations sometimes show up in attempts to define exactly what it is. On some understandings it is about dividing the properties into the safe, physical ones, and ones that seem more obviously supernatural, and seeing how far we get with the first ones. On another it is a matter of a defence of the catalogue of properties found in the natural sciences: limiting our acceptable properties to the ones which show up in current science, conservative extensions of current science, or perhaps ideal science.[10] There are ways of characterizing it as a continuity of nature thesis: the idea is that all the properties required to give an account of paradigm material things should suffice for an account of the mental. Finally there are methodological motivations: typically appealing not to the properties mentioned in science, but rather to the methods of science.

Whichever of these might be motivating, something like the methodological criteria seem to be clearly involved, no matter which materialist or immaterialist in the philosophical tradition of the last century we are talking about. One way to characterize these criteria is in terms of what a theory does with the raw 'data' of experience. One impulse is to use them, unanalysed, as either giving you information about the world, or as the building blocks of a theory. The other is to start with the experience itself, and take it to tell you things not (or not only) about the world, but about experience itself. The first impulse is the black boxes experience—it takes it either as a set of unanalysable simples for world building, or as what later came to be called 'transparent': telling you about other things, never about itself. The second takes experience itself to have intrinsic, maybe essential properties, which are revealed to an agent through having the experience. Experience on this understanding tells you about itself as much as about the world.

The latter impulse is likely to lead in various ways to versions of dualism (or at least of dual aspect theories or Russellian monism). For example, one contemporary way to characterize this impulse is in terms of the idea of the so-called doctrine of revelation. Plausibly, accepting this makes it very hard to resist neo-Dualist arguments like Chalmers' conceivability argument (which in fact pre-dates Chalmers at least as far back as Keith Campbell's 'imitation man' argument in 1970) or Jackson's knowledge argument (Campbell 1970; Jackson 1986).

---

[9] Armstrong (1968a, xi). See also Armstrong (1989, 11; 1993, xvi).

[10] This presents a trilemma for defining physicalism sometimes known as 'Hempel's' dilemma from a similar dilemma he proposed. See Hempel (1969).

But from the early twentieth century the other impulse was also strong. Resisting the phenomenologists' urge to introspect into details of experience as a basis for philosophizing, the positivists simply took experience as the beginning, not the end, of a theory. Experience is not about itself: experiences are observations, a word which both evokes the respectability of the empirical sciences, and which makes it clear that experiences are not about themselves. Observations are about something other. In his most idealist mode, Carnap takes experiences as the basic data from which the world can be constructed. Of course, this Carnap—the Carnap of the *Aufbau*—perhaps does not count as a materialist, though he certainly thought of himself as a naturalist, and it is doubtful that the characterization of the view as idealist is really correct. But the impulse that rapidly leads to materialism (and did so among some of his own person-stages, as well as his circle) is there. Experience is not something whose intrinsic nature is to be scrutinized by experience, its data are to be used in constructing a theory.[11] And the baroqueness of the theory so constructed leads to the idea that another way that experiences can be not about themselves is not to be primitives in a combinational theory, but rather to tell us directly about the world. They are observations.

But if all that is methodologically acceptable are observations (of material things, such as behaviours), then our world-building inherits much of what is problematic about Carnap's use of experience. We just have the tools of logic to recombine the things given in experience: this time we have behaviours and whatever else can be directly observed and recombined, rather than unanalysed experiences. But it does not tell us anything about that which lies *beyond* those observations. Instead it gives us *composites* of the things observed. This, perhaps, is at the heart of the many objections to behaviourism in all its philosophical forms.

What it takes to make progress is the idea of an explanatory theory which defines *new* properties in terms of the roles they play in explaining the observations. Such a theory solves the problem of how to infer beyond observations and also solves the problem of the apparent circularity involved in defining things in terms of their theoretical roles. To illustrate, consider a bunch of observations: there are instances of repulsion between things apparently of the same kind, there are things of different kinds spinning around in orbit, the things of a different kind appear to be attracted to each other. This is the familiar (and of course false) Bohmian early theory of the atom.

---

[11] Something like this might be a reconstruction of Feigl's approach, given in Feigl (1971).

How might we explain these observations? Perhaps there are three kinds of primitives: electrons, protons, and neutrons. Neutrons have no attachment to anything, electrons repel each other and are attracted to protons, protons are repelled by each other (though held together by a mysterious force in the nucleus) but are attracted to electrons, and so on. The complete picture gets its vindication by inference to the best explanation: it explains the observations. But what happens if we try to define these elements separately? For instance, maybe this is a sketch of the right definition:

> Electrons = def things attracted to protons and repelled by things of their own kind, etc.

But then protons appear in the definition of electrons. Fine; let us substitute in the definition of protons into the definitions of electrons. What is that? Perhaps it's something like this:

> Protons = def things repelled by themselves, but attracted to electrons.

But if we substitute that in the definition of electrons, we get:

> Electrons = def things attracted to things attracted to electrons…

It seems as though electrons are defined in terms of themselves in a way which might be viciously circular.

Frank Ramsey's insight was that the definition can be done in a holistic but uncircular way by existentially quantifying over the 'new' theoretical entities and saying that they are the things which play roles in the theories given, perhaps, by observations.

$$(x_1)(x_2)((x_1 \text{ A } x_2),(x_1 \text{ R } x_1),(x_2 \text{ R } x_2),(x_1 \text{ O } x_2)\ldots)$$

$$1 \quad 2 \quad 1 \qquad 2 \qquad 1 \qquad 1 \qquad 2 \quad 2 \qquad 1 \qquad 2$$

So what we would be saying is something like this. There are two kinds of things. Things of the first kind are attracted to things of the second, and vice versa. Things of the first kind are repelled by things of the same kind. Things of the second kind are also repelled. Things of the first kind orbit around things of the second kind. Then it only requires us to give a name to these things; the first kind are called electrons, the second protons.

To move from this crude account of atoms to the mind is of course no mere exercise for the reader, but perhaps its rough shape can easily be grasped. What are beliefs and desires, and how do they relate to behaviour? Well, there are two kinds of things. The first kind responds to the world, and changes an internal state. The second kind combines with that changed internal state and, depending on its own nature and the internal state it combines with, produces a behaviour. The first kind of things is called a belief. The second is called a desire. Such an account is at the heart of Lewis's classic paper on theoretical terms, and what made his own work on functionalism in his seminal papers 'An Argument for the Identity Theory' (Lewis 1966) and 'Psychophysical and Theoretical Identifications' (Lewis 1972) possible.

Of course so far we have only talked about explanation in general. But the immense richness of Armstrong's *Materialist Theory of the Mind* is his insistence that it be *causal* theories we use. This allows detailed hypotheses about all manner of mental states in terms, ultimately, of what they are apt to cause. We do not believe that anywhere in the literature since *MTM* has there been any attempt to provide such a rich and detailed account of mind in these terms, covering so many kinds of mental states (Lewis himself, and so many others including one of the current authors, concentrated on belief and desire in the hope that other mental states might be defined in terms of them).

It is our belief that *MTM* is something of a high-water mark. Yes there have been developments, but in our fallible opinion all the attempts to significantly change materialism have been retrograde. But perhaps we should conclude with a few words about the kind of target that a theory like this presents to neo-dualists and other immaterialists. A theory like this starts with experience, and in some cases, it ends with experience. It starts with experience understood as observation: transparently showing us things about the world. Experience shows us many things, but among them are behaviours, and we use a causal theory to infer the existence of states that play explanatory and causal roles in creating those behaviours. Some of those things are the very kind of thing we begin with: experiences. The materialist says they are the very same things, now understood better. The immaterialist says that this identity is hard to demonstrate. If experience has features that we are in touch with that are intrinsic to it, then this causal theory must leave those out, as it defines things not in terms of any such features, but rather in terms of typical causal roles: the effects it has on our instruments or ultimately our minds. In the metaphysics of mind, that has been the important battle after Armstrong.

References

Armstrong, D. M. (1960) *Berkeley's Theory of Vision*, Melbourne: Melbourne University Press.

Armstrong, D. M. (1961) *Perception and the Physical World*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1962) *Bodily Sensations*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1966) 'The Nature of Mind', *Arts* (Proceedings of The Sydney University Arts Association), 3, 37–48. Reprinted in Armstrong (1980), 1–15.

Armstrong, D. M. (1968a) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul, 2nd edition, London: Routledge, 1993.

Armstrong, D. M. (1968b) 'The Headless Woman Illusion and the Defence of Materialism', *Analysis*, 29: 48–9.

Armstrong, D. M. (1969) 'Dispositions Are Causes', *Analysis*, 30: 23–6.

Armstrong, D. M. (1973) 'Epistemological Foundations for a Materialist Theory of the Mind', *Philosophy of Science*, 40: 178–93.

Armstrong, D. M. (1977a) 'The Causal Theory of the Mind', *Neue Hefte für Philosophie*, 11, 82–95. Reprinted in Armstrong (1980), 16–31.

Armstrong, D. M. (1977b) 'Naturalism, Materialism and First Philosophy', in *Ist systematische Philosophie möglich?*. Proceedings of the Stuttgarter Hegel-Kongress 1975, edited by Dieter Henrich, Bonn: Bouvier Verlag Herbert Grundmann, 411–25. Reprinted in Armstrong (1980), 149–65.

Armstrong, D. M. (1980) *The Nature of Mind and Other Essays*, St Lucia: University of Queensland Press.

Armstrong, D. M. (1984) 'Self-Profile', in *D. M. Armstrong, Profiles*, vol. 4, edited by Radu J. Bogdan, Dordrecht: D. Reidel, 3–51.

Armstrong, D. M. (1989) 'C. B. Martin, Counterfactuals, Causality, and Conditionals', in *Cause, Mind, and Reality, Essays Honoring C. B. Martin*, edited by John Heil, Dordrecht: Kluwer Academic Publishers, 7–15.

Armstrong, D. M. (1993) 'Preface', in *A Materialist Theory of the Mind*, 2nd edition, London: Routledge, xiii–xxiii.

Armstrong, D. M. (1999) *The Mind-Body Problem: An Opinionated Introduction*, Boulder, CO: Westview Press.

Armstrong, D. M. (2010) *Sketch for a Systematic Metaphysics*, Oxford: Oxford University Press.

Campbell, Keith (1970) *Body and Mind*, New York: Anchor Books.

Chalmers, David J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press.

Feigl, Herbert (1950) 'The Mind-Body Problem in the Development of Logical Empiricism', *Revue Internationale de Philosophie*, 4: 64–83.

Feigl, Herbert (1958) 'The "Mental" and the "Physical"', *Minnesota Studies in the Philosophy of Science*, vol. 2, Minneapolis, MN: University of Minnesota Press, 340–497.

Feigl, Herbert (1971) 'Some Crucial Issues of Mind-Body Monism', *Synthese*, 22: 295–312.

Hempel, Carl G. (1969) 'Reduction: Ontological and Linguistic Facets', in *Philosophy, Science and Method: Essays in Honor of Ernest Nagel*, edited by Sidney Morgenbesser, Patrick Suppes, and Morton White, New York: St Martin's Press, 179–99.

Jackson, Frank (1986) 'What Mary Didn't Know', *Journal of Philosophy*, 83: 291–5.

Jackson, Frank, Robert Pargetter, and Elizabeth W. Prior (1982) 'Functionalism and Type-type Identity Theories', *Philosophical Studies*, 42: 209–25.

Lewis, David (1966) 'An Argument for the Identity Theory', *Journal of Philosophy*, 63: 17–25.

Lewis, David (1972) 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy*, 50: 249–58.

Lewis, David (1980) 'Mad Pain and Martian Pain', in *Readings in Philosophy of Psychology*, vol. 1, edited by Ned Block, D. Terence Langendoen, and Jerrold J. Katz, Cambridge, MA: Harvard University Press, 216–22.

Medlin, Brian (1969) 'Materialism and the Argument from Distinct Existences', in *The Business of Reason*, edited by J. J. Macintosh and S. Coval, London: Routledge & Kegan Paul, 168–85.

Ramsey, Frank P. (1978) 'Theories', in *Foundations: Essays Philosophy, Logic, Mathematics and Economics*, edited by D. H. Mellor, London: Routledge & Kegan Paul, 101–25. First published in 1929.

Ryle, Gilbert (1949) *The Concept of Mind*, London: Hutchinson.

Ryle, Gilbert (1951) 'Feelings', *The Philosophical Quarterly*, 1: 193–205.

Squires, Roger (1968) 'Are Dispositions Causes?', *Analysis*, 29: 45–7.

2

Lewis and the Identity Theory

D. M. Armstrong

I discuss here David Lewis's important paper 'An Argument for the Identity Theory' published in the *Journal of Philosophy* in 1966. It was his first published paper and the affiliation given for the paper is Harvard University. That means that it was written when he was still a graduate student. And indeed there exists a term paper of David's dated 1 May 65 called 'How to establish the Identity Theory' on which David has written: 'term paper for Donald Williams' course *Metaphysics* Spring 1965'. It isn't too different from the published paper. (I am very grateful to Stephanie Lewis for making a copy of the term paper available to me together with Williams's comments and a letter David wrote to Smart at that time.)

With Stephanie's help I have been able to fill in the background still further. During the academic year 1959–60, David took a year off from his undergraduate studies at Swarthmore to accompany his parents to Oxford during his father's sabbatical year from Oberlin College. He was planning to take his major in chemistry, but was already interested in philosophy. Arriving at Oxford, which at that time believed it was leading the world in a philosophical revolution, he decided to spend the year studying philosophy. He was enrolled in St Catherine's Society—his tutor, by the way, was Iris Murdoch. (It will be seen, also by the way, that the possible world in which David Lewis did not become a philosopher may not be a very distant one.)

One thing that David took from Oxford was an enduring respect for conceptual analysis. I think that this respect becomes particularly clear in his work on causation. His counterfactual account of causation, which started off nice and simple, became so exceedingly complex just because he wanted an account that would cover all the twists and turns of the ordinary concept of cause. But although causation comes into the story here, I will not be further concerned with Lewis's *theory* of causation. In his philosophy of mind, he does indeed make central use of the concept of causation. But he treats it there as a primitive.

But one concept that certainly caught his attention in Oxford was the concept of mind. Ryle's book *The Concept of Mind* had appeared in 1949. By 1959, in Oxford, few had advanced beyond it.[1] A lot of people were dissatisfied with Ryle's

---

[1] This was still true in 1962, when Brian Medlin and I gave a seminar in Oxford defending the Identity theory. With the honourable exception of Anthony Quinton, the philosophers there gave us little beyond the argument from the raised eyebrow (an Oxford variant of the incredulous stare).

D. M. Armstrong, *Lewis and the Identity Theory* In: *Armstrong's Materialist Theory of Mind.* Edited by: Peter R. Anstey and David Braddon-Mitchell, Oxford University Press. © D. M. Armstrong 2022. DOI: 10.1093/oso/9780192843722.003.0003

account, but few there seemed to have a clue about how to move to some larger perspective that might be more fruitful in thinking about the mind. In Oxford, one still started from Ryle's book. It is true that, in 1956, U. T. Place, himself originally an Oxford philosophy undergraduate, had published a paper, 'Is Consciousness a Brain Process?'. But this paper appeared in *The British Journal of Psychology* and nobody in philosophy appears to have taken much notice of it until Place converted Jack Smart to the theory (a process which occurred in Adelaide, South Australia). Smart then advanced the cause with his 'Sensations and Brain Processes', published in 1959 in *The Philosophical Review* and read very widely. It is interesting to note that Lewis in his article makes no mention of Place's paper, and I suppose that he did not read it until well after that year at Oxford. Probably he did not see Smart's paper either at that time. But he was a physicalist already, something derived, I suppose, from his early scientific bent. If you are a physicalist and if the only available current philosophical work on the mind that is non-dualist seems to be Ryle, then you will lean to Ryle, reading him as a philosophical behaviourist. That, indeed, was the path that was taken by Place, by Smart, and by myself. We have the following testimony that seems to show that Lewis, considerably younger and abler than the rest of us, took the same path: 'I am an ex-Rylean, and I retain some part of the Rylean legacy' (Lewis 1994, 412). It was Place, in this gang of four, who led the way to a *central-state* materialism.[2]

In the fall of 1963, however, Smart was giving a graduate seminar on Space and Time as a visitor to Harvard. Lewis attended that class and saw a great deal of Smart. Then, if not before, he would have seen the Smart paper. Plenty of time for him to write that term paper for Donald Williams which, only lightly revised, became his first publication.

One should briefly mention one respect in which he stayed with Place and Smart. He confined the identity he was arguing for to, as he put it, *experiences*. In later work, however, he tacitly moves beyond this to mental states generally, for instance, in 'Mad Pain and Martian Pain' (see p. 124 of that paper). But that said, Lewis immediately introduces a great improvement into the Identity theory, though I now need to back up a little to make clear the nature of the improvement.

It had been obvious since Smart's paper that if the Identity theory was to succeed, our concept of mind was going [to] have to contain a *topic neutral* component. For what, after all, was the stock objection that the vast majority of philosophers made to the thesis that mental states were states of the brain when they first heard it suggested by Place and Smart? It was that our mental concepts long pre-dated any knowledge that there was a link between the mental and our brain. What then, they asked, was there about these concepts of mind and brain that permitted this identification? Or [to] put the problem in a somewhat different way. We have direct acquaintance with mental states, at least in our own case. (The fatal problem for philosophical behaviourism.) We do not experience them as brain states. What then do we experience them *as*? And how can what we experience be identified with states of the brain?

---

[2] A useful phrase introduced by Herbert Feigl, whose physicalism, however, was not as thoroughgoing as that defended by Lewis and the rest of the gang of four. See his 1958, reprinted with an addition in 1967 (Feigl 1967).

I don't think that Place got very far in facing up to this problem. Smart got a bit further. What Smart said (taking the particular case of an after-image) was:

> When a person says, 'I see a yellowish-orange after-image', he is saying something like this: '*There is something going on which is like what is going on when* I have my eyes open, am awake, and there is an orange illuminated in good light in front of me, that is, when I really see an orange'. (Smart 1959, 149)

The great point of the 'There is something going on which is like what is going on' is that it is topic neutral (a phrase Smart took from Ryle). It provides a blank into which brain processes (*or*, of course, something that is not a brain process) can be plugged in. This, I think, was Smart's great contribution to the development of the Identity theory.

David, however, sums up Smart's answer as: '... avowals of experience are, in effect, of the form "What is going on in me is like what is going on in me when ..." followed by specification of typical stimuli for, or responses to, the experience' (Lewis 1966, 20). That is already a considerable massaging of Smart's answer, changing it to give the Lewisian answer. Stimulus and response are *causal* concepts, and Smart appears to make no appeal to causation. And even if what he says can be read as meant to refer to *stimuli*, responses get no mention at all. Yet bringing in responses is an enormous change and a much-needed improvement. Furthermore, Lewis immediately goes on to say:

> I wish to add explicitly that ... [the specification of typical stimuli for, or responses to, the experience] may be an elaborate logical compound of clauses if necessary; that ... [the specification] must specify typical causes or effects of the experience, not mere accompaniments; that these typical causes and effects may include other experiences; and that the formula does not apply only to first-person reports of experience. (Lewis 1966, 20)

With his genius for putting things in a brief and pregnant way, he sums up his view by saying that the concept of a mental state (let us now go beyond experiences) is the concept of that which plays a certain causal role. I am very envious of this beautiful formula because I have the honour, also starting from Smart's insight, to have arrived at the same time at much the same account as Lewis's of the concept of a mental state (Armstrong 1966; 1968). But I failed to see how simply and neatly it could be presented.

Once causation has been brought in, the Identity theory can have good hope of becoming an unpuzzling theory. Let the mental concepts be purely causal concepts, concepts of states and processes not otherwise specified except by their different, complex, and interactive causal roles. But let them be anchored, as it were, to human and other animal bodies, anchored as states and processes that lie causally between the stimulus and the response.[3]

This view of the mental is at least very much more plausible than the analyses of the mental offered by philosophical behaviourism. For myself, I still think it is on the right track, though I am less sure than I once was that the account is no more than a conceptual analysis. (David, though, never wavered in his faith in conceptual analysis here.)

But suppose you accept this. Then, of course, it becomes immensely plausible that what these causally defined states and processes are in their own nature, what these things are that do the causal job between stimulus and response, are purely physical states and processes in the brain. That is why Lewis calls his paper 'An *Argument* for the Identity Theory'. That is why he rejects the view that merely Occamist considerations are what should decide for the Identity theory and against Dualism, a line to which Smart, in particular, to an extent inclined. Once the doctrine of 'the mental as essentially causal' has been accepted, then since we already know that the brain has an important causal role in our reacting to stimuli as we do, it will be a natural hypothesis that the physical brain is the bearer of all the mind's causal capacities. Occamist considerations, one would suppose, only come in if it is suggested that perhaps the brain does not do the whole job.

It seems clear, furthermore, that, as Lewis says, that here we have a *contingent* identity. That which instantiates the mental causal role might not have instantiated that role, and something else, something spiritual perhaps, might have done this causing instead. In his 'Mad Pain and Martian Pain' (Lewis 1980), which is a beautiful coda to the original paper, he draws the consequence that the word 'pain', for example, is a non-rigid designator. It is possible to have what he calls Martian pain, where the concept of pain remains the same but pain is not a brain process. It is possible to have what he calls Mad pain where the brain process is the same as our pain, but the causal role of the brain process is quite different. The Identity theory has the flexibility to allow such possibilities.

So I think he left the Identity theory in pretty good shape. You can, of course, quarrel with the purely causal analysis of the mental concepts. But if you accept that analysis, it will be hard to reject the next step.

---

[3] An important and clarifying model here is Brian Medlin's comparison of mental states and processes to the genes when they were conceived of as bare, because unknown, factors. They are, of course, causal factors, factors that spring from the genes of the parental organisms ('stimulus') and that in turn produce the macroscopic characteristics of the inheriting organism ('response'). See Medlin (1969), a paper first read in Oxford in 1962.

References

Armstrong, D. M. (1966) 'The Nature of Mind', *Arts* (Proceedings of The Sydney University Arts Association), 3: 37–48. Reprinted in Armstrong (1980), 1–15.

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1980) *The Nature of Mind and Other Essays*, St Lucia: University of Queensland Press.

Feigl, Herbert (1967) *The 'Mental' and the 'Physical': The Essay and a Postscript*, Minneapolis, MN: University of Minnesota Press. The essay originally appeared in *Minnesota Studies in the Philosophy of Science*, 2 (1958): 370–497.

Lewis, David (1966) 'An Argument for the Identity Theory', *Journal of Philosophy*, 63: 17–25. Reprinted, with additional material, in Lewis (1983), 99–107.

Lewis, David (1980) 'Mad Pain and Martian Pain', in *Readings in the Philosophy of Psychology*, vol. 1, ed. Ned Block. Cambridge, MA: Harvard University Press, 216–22. Reprinted in Lewis (1983), 122–32.

Lewis, David (1983) *Philosophical Papers*, vol. 1, New York: Oxford University Press.

Lewis, David (1994) 'Reduction of Mind', in *A Companion to the Philosophy of Mind*, edited by Samuel Guttenplan, Oxford: Basil Blackwell, 412–31.

Medlin, Brian (1969) 'Materialism and the Argument from Distinct Existences', in *The Business of Reason*, edited by J. J. Macintosh and S. Coval, London: Routledge & Kegan Paul, 168–85.

Place, U. T. (1956) 'Is Consciousness a Brain Process?', *British Journal of Psychology*, 47: 44–50.

Ryle, Gilbert (1949) *The Concept of Mind*, London: Hutchinson.

Smart, J. J. C. (1959) 'Sensations and Brain Processes', *The Philosophical Review*, 68: 141–56.

3

The Two Davids and Australian Materialism

A. R. J. Fisher

1. Introduction

The rise of the identity theory in the 1950s and 1960s was due in part to a new version of materialism that came out of Australia—aptly called 'Australian materialism'. It shifted the debate from dated criticism concerning meaning and knowledge to an appreciation for the metaphysical concepts that underpin philosophical theorizing about the mind. At the same time it gave science a direct role in telling us what the mind is. Australian materialists offered an account of the *nature* of mind. Their version of the identity theory was an ontological thesis. It flew in the face of evolving positivist doctrine and ordinary language dismissals of the mind-body problem. In this respect, Australian materialism was one of the first theories in mid-twentieth-century analytic philosophy that was overtly metaphysical. On top of that, it encapsulated a certain tendency, as S. A. Grave puts it (1976, 23), that helped to define Australian philosophy: philosophy concerns what there is; it is not purely about conceptual matters; and it should put forth a worldview, i.e., a systematic account of reality, and one that is continuous with science.

The main figures in the Australian materialist tradition are U. T. Place, J. J. C. Smart, D. M. Armstrong, and David Lewis. Place (1956) set things in motion by arguing that the statement 'consciousness is a brain process' is an empirically reasonable hypothesis. Smart (1959) defended the identity theory against then current objections and proposed a topic-neutral account of sensation-reports. In *A Materialist Theory of the Mind* (hereafter *MTM*), Armstrong completes the argument thus: a general conceptual analysis of *mind* can be systematized and rendered compatible with the identity theory. In America, Lewis argued for the identity theory—first, by generalizing Smart's topic-neutral account of sensation-reports (Lewis 1966), and, second, by exploiting Frank Ramsey's method of eliminating theoretical terms in the case of folk psychology (Lewis 1972). Both Armstrong and Lewis (independently and almost simultaneously) proposed that the concept of a mental state is functionally defined in terms of certain input and output clauses and pointed out that science has discovered what satisfies these functional definitions. Thus, pain is whatever

A. R. J. Fisher, *The Two Davids and Australian Materialism* In: *Armstrong's Materialist Theory of Mind.* Edited by: Peter R. Anstey and David Braddon-Mitchell, Oxford University Press. © A. R. J. Fisher 2022.
DOI: 10.1093/oso/9780192843722.003.0004

fills the pain-role; science tells us that C-fibres firing fills the pain-role; hence pain = C-fibres firing.

According to standard histories of analytic philosophy of mind, the identity theory burst onto the scene in reaction to logical behaviourism. In this story,[1] logical behaviourism is the view that the meanings of mental terms or sentences are translatable into the meanings of physical terms or sentences about behaviour or behavioural dispositions. This view is a semantic thesis and the analyses offered are conceptual and knowable a priori. Rudolf Carnap, Carl Hempel, and Gilbert Ryle, so the story goes, held this position. Then the identity theorists came along in the 1950s and 1960s, namely, Herbert Feigl, Place, and Smart, and proposed that the purported identity between the mental and the physical is factual, not conceptual, and knowable a posteriori. Our theory of mind should not be fashioned after linguistic definitions of terms but after 'empirical discoveries of identities in science', as John Searle claims when telling the story (Searle 2004, 55). Thus, logical behaviourism was supplanted by the identity theory (see Kim 1998, 2).

Sean Crawford (2013) has recently shown that this story has many holes. He argues that logical behaviourism in the positivist tradition has been misunderstood. Carnap's behaviourism, in particular, is not the view that translations between mental and physical statements are analytic. Rather, they are *synthetic*; more precisely, they are material, extensional equivalences between mental and physical statements (these statements are P-equivalent, to use Carnap's terminology). Crawford further argues that Feigl's (1958) identity theory stems from the pre-positivistic stage of Moritz Schlick, who was the first twentieth-century identity theorist (Crawford 2013, 643–5). Crawford explains that the real change in the 1950s and 1960s as regards the mind-body problem 'was simply the rejection of the earlier view, shared by both logical positivists and ordinary language philosophers, that it was a pseudo-problem' (Crawford 2013, 639). However, Australian materialism and its distinctive history should be looked at more closely.

For one thing, Feigl's identity theory is quite different from Place's and Smart's. Feigl formulates the view within a critical realism that regards the data of immediate experience as the 'ultimate testing ground' (Feigl 1975, 22). The buck stops with the phenomenally given, so to speak. What is more, Feigl embraces Schlick's epistemological parallelism, thus placing himself within the psychophysical parallelist tradition (Heidelberger 2001; Stubenberg 1997). On this view, there is one reality designated by concepts of two wholly distinct conceptual systems: a physical (intersubjective) conceptual system that consists of physical concepts and a psychological (subjective) conceptual system that consists of psychological concepts. When it comes to qualia, Feigl thinks they are structural (i.e., complex) properties that we know by acquaintance in virtue of having them. In our introspective description of phenomenal properties 'we deal with their *structural* features' (Feigl 1975, 28, his italics). Crucially, in specifying what is identical with what, Feigl says: '[t]he identity is that of the *structure* of the phenomenally given with the *structure* of certain global aspects (Gestalten) of the processes in the cerebral cortex' (1975, 31, his italics). Because Feigl places more importance on the phenomenal side of the mind-brain identity, some philosophers have labelled his view the 'qualitative identity theory' (Pepper 1975, 37) and a 'mentalistic form of the identity theory' (Crawford 2013, 646).

---

[1] I borrow this characterization of the standard story from Crawford (2013, 621–2).

Smart, on the other hand, advocates a physicalistic form of the identity theory founded on an unabashed scientific realism, an uncompromising reductionism, and Ockham's razor. He in effect makes the opposite moves to Feigl. He prioritizes the physical over the mental, rejects nomological danglers, closes every door to emergentism, and puts epistemology second. He also denies that there are any qualia, let alone assigns them a fundamental epistemic role in his theory. Furthermore, there are differences of tradition and historical context. Australian materialism does not develop out of logical positivism or logical empiricism. Place's 'Is Consciousness a Brain Process?' is actually a continuation of an earlier paper wherein he accepts Ryle's behaviourism (Place 1954). Although Ryle is misinterpreted on this point, he does not believe that *all* mental concepts are subject to a behaviourist treatment. He thought that his behaviourist treatment did not apply to the concept of having a sensation, and he recognized this as a shortcoming (Ryle 1949, Chapter 7, Section 1). Some philosophers did read Ryle correctly. In addition to Place, Iris Murdoch and Julia Tanney interpret Ryle in this way (Murdoch 1956, 37; Tanney 2009, xxviii).

Place's disagreement with Ryle was specifically about 'heed' concepts (such as attending to, observing, being conscious of something). Place thought they had to refer to some internal state (Place 1954, 254). He thus restricted the scope of behaviourism: the recalcitrant sensory concepts gave way to an identification of sensations with brain processes. In one respect he can be seen as improving Ryle's position. In a more radical vein, Smart initially held a full-blown (unrestricted) behaviourism about *all* mental concepts, convinced that Ryle's position could be extended to cover sensory concepts. In a largely forgotten article he explicitly states that he prefers the Wittgensteinian expressivist account of sensation-reports to dualism and the identity theory (Smart 1957, 76). But due to three-way discussions with Charlie Martin and Place, Smart came around to Place's restricted behaviourism (Smart 1959) and then the identity theory (Smart 1967).[2]

---

[2] C. F. Presley says: 'Smart, who describes himself as having previously been "very Rylean", became converted to Place's view when, as visiting Professor at Princeton in 1957, he was conducting a graduate class on Wittgenstein and Ryle. In November of that year, he read a paper at Cornell in which he

The conceptual breakthrough for Place and Smart was the realization that the identity theory is not a translation thesis between sensation statements and brain-process statements. It is an ontic thesis. At the time, translation theses were popular. Phenomenalism was typically formulated as the view that statements about material objects are translatable into statements about sense-data (as this was the most viable way to make sense of the proposal that material objects are logical constructions out of sense-data) (see Ayer 1936, Chapter 3). It was eventually recognized that phenomenalism as formulated is false, because it is impossible to translate material-object statements into sense-data statements in the relevant sense (Wisdom 1953, 139).[3] Even Smart acknowledged (using the then typical example of some nation and the people of that nation) that it was commonplace that 'no translation from nation statements to person statements is possible' (Smart 1955, 51). Some philosophers took the failure of translatability to imply the irreducible existence of the entities of which the irreducible statements are about (e.g., Mandelbaum 1955, 310). But Place and Smart (and probably Martin) realized that even though it is impossible to translate X-statements in terms of Y-statements, it does not follow that the Xs are not identical with the Ys. Why think that statements and facts about translation possess such ontic power? Phenomenalism is about material objects, not about translations of statements concerning them. Analogously, the identity theory concerns the mind, not statements about it. Thus began one ontological turn away from understanding things in linguistic form.[4]

Place and Smart arrived at their formulation of the identity theory by working through problems with phenomenalism and drawing analogies with the identity theory. As such, their theory of mind develops out of ordinary language philosophy and British analytic philosophy, not positivist behaviourism. Therefore, there is no single origin of the identity theory in analytic philosophy of mind.[5] There are separate lines of influence running through differing versions of the identity theory and distinct reactions to logical behaviourism, which in turn afford separate explanations of the transition from logical behaviourism to the identity theory. The transition from logical behaviourism to the identity theory in the positivist tradition might plausibly be interpreted as a change from pseudo-problem (Carnap, positivistic Feigl) to genuine problem (post-positivistic Feigl). But the

---

defended Place's view' (Presley 1967, xi). No doubt that paper was 'Sensations and Brain Processes'. It should be noted that Smart had not read Feigl until 1957 (Presley 1967, xiii). In these discussions, Martin held some version of property dualism.

[3] The relevant sense was probably a conceptual kind of necessity.

[4] For a historical overview of Place's and Smart's theory, see Grave (1984, Chapter 6).

[5] Charles Wolfe implies that the genesis of Australian materialism is found in the Vienna Circle (and their criticism of vitalism) (Wolfe 2015, Chapter 7, esp. 93). Such an interpretation relates distinct traditions in the wrong way. Unfortunately, it is common for Place, Smart, and Feigl to be grouped together when philosophers present the identity theory of the 1950s, as if temporal proximity implies some intellectual connection or influence in doctrine. For a recent example, see McLaughlin and Planer (2014).

transition for Australian materialists is slightly more subtle than this. They always regarded the mind-body problem as genuine and either from *their own* occupation of a behaviourist standpoint—stemming from the ordinary language philosophy tradition—ended up as identity theorists (as in the case of Place, Smart, and Lewis) or from a serious consideration of behaviourism as a form of materialism rejected it in favour of the identity theory (as in the case of Armstrong).

Armstrong and Lewis (the two Davids) went on to become highly influential philosophers in late analytic philosophy. Their version of materialism came to be known as analytical or common-sense functionalism and it contributed to the rise of functionalism in analytic philosophy of mind. There is no in-depth account of their path from behaviourism to materialism in the literature. Such an account is warranted and required to understand the origins and development of Australian materialism. Moreover, their respective theories of mind shaped their philosophical method and other views in philosophy. Thus, an account of their early thought on the philosophy of mind would assist in understanding how it relates to other aspects of their philosophy.

In what follows, I give an account of Armstrong's thought leading up to and including *MTM*, identifying aspects of *MTM* that are continuous with his earlier thought and that shape his later philosophy and philosophical method (§2). I analyse the origins and development of Lewis's identity theory along similar lines (§3). I conclude by reflecting on where we should locate Australian materialism in the history of philosophy, broadly speaking (§4).[6]

2. Armstrong and *MTM*

Armstrong began his philosophical training at the University of Sydney in 1947. By then, John Anderson had been the Challis Professor for two decades, exerting a powerful influence not just on Philosophy but also on Psychology, English, the university, and the wider Sydney community (see Franklin 2003). Anderson was a staunch defender of metaphysics, metaphysical realism, metaphysical naturalism, and direct realism about perception. Central aspects of Anderson's metaphysics and philosophy of mind are developed in reaction to Samuel Alexander's realist metaphysical system (Alexander 1920; for discussion, see Fisher 2015). Throughout his career Anderson delivered many lectures on improvements to Alexander's metaphysics (e.g., Anderson 2005).[7] As an undergraduate, Armstrong attended Anderson's 1949–50 lectures (Anderson 2007).[8] As noted by Armstrong (e.g., 1993b, 169) and others (Bacon, Campbell, and Reinhardt 1993, vii), Anderson was crucial in the early formation of Armstrong's thought. He set Armstrong on the path of realism and metaphysical naturalism, instilled in Armstrong the belief that metaphysics is not a fruitless or fraudulent enterprise, and imparted an interest in the history of philosophy.

---

[6] Although I draw heavily from Lewis's letters in §3, I do not take his correspondence with Armstrong into account. It falls outside the scope of this chapter because both of them had adopted the identity theory by the time their correspondence began in 1968.

[7] Anderson lectured on Alexander as early as October 1924—at the University of Edinburgh (John Anderson Papers, P.42, Series 1, box 2, item 025. Sydney University Archive).

In 1952, Armstrong was awarded a Sydney University travelling fellowship. He left for Oxford to complete a BPhil at Exeter College. Despite being surrounded by the often narrow focus on words and on what one might 'say', Armstrong found books that aligned with and reinforced his extant interests and realist tendencies. Luckily, he had to look no further than his supervisor Henry H. Price, who was a product of the Cook Wilson brand of early analytic philosophy that flourished at Oxford before the Second World War. In this period Price wrote two books relevant to Armstrong's early development: *Perception* (Price 1932), in which Price defends a form of the sense-datum theory, and *Hume's Theory of the External World* (Price 1940), which is an account of Hume's theory of perception based on a textual analysis of the *Treatise*.

It is clear that Armstrong was affected by Price's interests and his work. The influence, to be precise, is due more to Armstrong's reading of Price's work than philosophical conversations they might have had while Armstrong was at Oxford for his BPhil. As Armstrong later reports, the two of them 'never seemed to get around much to discussing philosophy' (Armstrong 1984, 11). Armstrong's first book, *Berkeley's Theory of Vision* (Armstrong 1960) is a critical commentary on Berkeley's *New Theory of Vision*, very much in line with Price's method in history of philosophy (Price 1940, 3) and his treatment of Hume. Price even gave Armstrong substantial comments on a draft typescript of the book (Armstrong 1984, 11), and as one reviewer remarks, it is 'pleasantly reminiscent of Professor Price's classic on Hume's theory of perception' (O'Connor 1963, 472). In this book, after a careful exposition of Berkeley's theory of perception, Armstrong argues that Berkeley is wrong to maintain that distance is not immediately seen. A crucial part of his criticism is a key premise he adopts from Price—the claim that visual sense-data are not two-dimensional; visual sense-data have depth (Price 1932, 245). Armstrong further argues: 'the whole situation, "colour-change preceded by temperature-change", is perceived by both sight and touch, and it makes no sense to ask which part of this relational situation is perceived by sight and which part by touch' (Armstrong 1960, 84). So, against Berkeley, visible and tangible extension are one and the same. Similarly, Price thought that visual and tactual sense-data can be spatially related to each other (Price 1932, 244–6).

---

[8] Armstrong's copy of Alexander's *Space, Time, and Deity* is dated '1949'. Thanks to Peter Anstey for allowing me to browse Armstrong's library.

Despite the obvious effect of Price, Anderson's influence was stronger with respect to first-order doctrine. Armstrong never accepted the sense-datum theory. Following Anderson, he endorsed direct realism. We see sympathies towards direct realism in his book on Berkeley (Armstrong 1960, 8–9, 92–3), not to mention two earlier articles that defend the view (Armstrong 1954; 1955). His second book, *Perception and the Physical World* (Armstrong 1961) carries the argument for direct realism forward.[9] His criticism of competing views is informed by his reading of Berkeley just as much as contemporaneous literature on phenomenalism. And his criticisms go to the metaphysical heart of the view under scrutiny. For instance, one of his objections against phenomenalism is that it cannot individuate two distinct minds existing at the same time without positing some ad hoc principle beyond our direct acquaintance (Armstrong 1961, 69–70).

During this same period Armstrong continued his research into philosophical psychology. Having taken an interest in Ryle's discussion on feelings (Ryle 1949, 84, 104; 1951) Armstrong wrote his third book, *Bodily Sensations* (Armstrong 1962). Such sensations as itches, tickles, pains, aches, etc. as well as sensations of heat and cold, are cases of bodily sensations. Ryle hints at the view that feeling-words designate bodily sensations but his remarks are somewhat evasive and boil down to linguistic recommendations about how uses of 'feel' are related. In contrast, Armstrong gives an account of the *nature* of bodily sensations.

Armstrong's thesis is that bodily sensations are sense-impressions of a certain sort.[10] Bodily sensations are not sensible qualities of parts of one's body. To illustrate, suppose I have a bodily sensation of heat. For Armstrong, it is an impression of heat. If I have a bodily sensation of pain, it is an impression of pain. So the sentence 'my hand feels sore' is akin to 'my hand feels hot' (Armstrong 1962, 59). While Armstrong is concerned with the nature of bodily sensations, it is part of his model of conceptual analysis that we give faithful accounts of how ordinary language ascribes certain features to bodily sensations. This is not ordinary language philosophy but serious conceptual analysis about the world. As Armstrong focuses more directly on ontological issues in his later work he moves away from couching the main issues in terms of conceptual analysis and translations among sentences. But at this point he remains close to Oxford (see Armstrong 1993a, the Preface to *MTM*).

---

[9] In the Preface of both books, he acknowledges his debt to Anderson's teaching (Armstrong 1960, Preface; Armstrong 1961, ix). In his later work on the metaphysics of properties, he dedicates *Universals and Scientific Realism* (1978) to Anderson and notes the influence of Anderson's ontology of facts (Armstrong 1997, 3–4). The force of the influence by the late 1970s is not as strong as his avowals make out. As A. J. Baker notes, despite the dedication in *Universals and Scientific Realism*, Armstrong does not discuss Anderson's theory of universals at length (Baker 1986, 62, n. 1). His recognition of any influence is distant and more of a tribute than a close reading and engagement with Anderson's work. But this does not make the connection any less significant. The tribute is to the first philosopher that showed him how to do metaphysics at a crucial developmental stage of his thought.

[10] Whatever sense-impressions turn out to be on one's preferred account of sense-impressions; his view is that sense-impressions are conscious acquirings of beliefs (Armstrong 1961, 128–32).

His first book contains important remarks about his conception of philosophy from his early period. When discussing the relevance of experimental evidence in theories of perception, he considers the objection that philosophy should butt out of such inquiries. So the objection goes: 'philosophy, . . . moves in the sphere of logical necessity: it is concerned with, or can only profitably be concerned with, conceptual, or at any rate non-empirical, inquiries. For the philosopher to seek to discover the nature of reality is a presumptuous encroachment on the domain of science' (Armstrong 1960, 61). Such an objection would have rolled off the tongue of a logical positivist or an 'analytical' philosopher who thought that philosophy's objective is to clarify the meaning of already known statements.

Armstrong's response is that experimental evidence, specifically evidence from experiments of cataract patients blind from birth and newly born animals, is indeterminate in the dispute about whether distance is immediately seen and whether the object seen is identical with the object touched. Both Berkeley and his opponent can explain away troublesome evidence against their view. Any scientific observation resides within a body of background assumptions, such that any meaningful results are anchored in that context. The domain of science and empirical facts in this domain will not decide the issue. Part of the reason for this, Armstrong says, is that the kind of question Berkeley is raising is a 'border-question' (Armstrong 1960, 65). It lies on the border between philosophy and science. The empirical observations that are relevant to such a border-line inquiry are the observations that are 'available to everybody' (Armstrong 1960). One defining feature of philosophical inquiry is 'the scrutiny of the obvious' (Armstrong 1960). This is not to say that scientific considerations are left out of philosophical inquiry. Given that the inquiry is on the border, both philosophy and science share in the pursuit of uncovering the truth about perception.

By the time Armstrong wrote *MTM*, his understanding of the connection between philosophy and science had changed. The change, in fact, occurred a couple of years earlier in 'The Nature of Mind' (Armstrong 1966), wherein he first defined the scientism that put him on the path to his materialist theory of mind. In *MTM*, science plays a much more central role in philosophical debate than in his book on Berkeley. In *MTM*, he puts forth 'the argument from the supremacy of physics' (*MTM*, 51). This argument suggests that scientific results give us inductive reasons to suppose that psychology will be reduced to physics. In defence of this argument, he considers a similar objection to the one he entertained in his book on Berkeley: philosophy is concerned with conceptual truths to the exclusion of scientific truths. His first reaction is to say that: 'philosophy is not solely conceptual analysis' (*MTM*, 51). His second response is that even if philosophy is solely concerned with conceptual analysis, the correct analysis of some concepts cannot be supported solely by a priori propositions. Ordinary experience and science can help with showing where the truth lies in conceptual analysis. He adds that scientific considerations should be given epistemic weight and points to the scientific success that came out of the early modern period. These successes advanced our philosophical understanding of the world as well. As a result, a major premise for materialism is that science has discovered that the central nervous system is the mediator between stimulus and response (*MTM*, 79). This directly informs our philosophical analysis of the concept of a mental state.

Smart's reductionism and scientific realism are two factors in Armstrong's shift towards a further privileging of science. This marks an important turning point in Armstrong's philosophy because it is here that we can locate the seeds of his a posteriori realism. On this approach, metaphysics answers questions about abstract, general concepts, such as property, object, ontological category, etc., while science discovers what there is in the sense that it discovers the specific contents of our given ontology. A priori metaphysics and a priori arguments derived from language do not determine what there is. Given Armstrong's ontology of immanent universals, science—not a priori or linguistic reasoning—discovers what universals there are. (Even here, the influence of Price can be detected. In *Thinking and Experience*, Price rejects the argument from meaning (for the existence of universals), asserting that what matters most in ontic reflections is what there is (Price 1953, 11).)

Armstrong's second and third books are important for understanding the development of his materialism too. In both works he appeals to causation in giving philosophical analyses of perception and bodily sensations respectively. In both cases he owes this insight to Martin, with whom he discussed these topics from 1956 to about 1962. According to Armstrong's theory of perception, sense-impressions are conscious acquirings of belief or dispositions to believe something is the case. Veridical perception is a case of acquiring knowledge; sensory illusion is a case of acquiring false belief. Martin pointed out to Armstrong that the fact that a sense-impression corresponds to reality is not enough for veridical perception. The missing ingredient is a causal relation that holds between the object in the external world and the sense-impression (Armstrong 1961, 125). This causal component is part of the concept of veridical perception (see also Armstrong 1989, 11–12).[11]

According to Armstrong's theory of bodily sensations, the bodily sensation of pain is a bodily sense-impression, as mentioned above. But it is a sense-impression that leads to attitudes in the individual (Armstrong 1962, 106–7). This suggests that the concept of pain involves two things: (1) having a bodily sense-impression; and (2) having an attitude to that sense-impression. The sentence 'my hand feels sore' is translated as 'I have a bodily sense-impression that invokes an attitude in me towards it'. If this sense-impression corresponds to reality, it is an instance of veridical perception. So the damage to the body is the cause of the sense-impression (Armstrong 1962, 110). It follows that the 'place' of the pain is the place where the pain is felt and is the cause of the bodily feeling/sensation.

---

[11] Martin had already put forth his causal theory of perception in *Religious Belief* (Martin 1959) and went on to articulate a causal theory of remembering; see Martin and Deutscher (1966).

This emphasis on causation had a lasting impact on Armstrong since the idea of applying causation to philosophical analyses of concepts led him in part to his analysis of the concept of a mental state—that which (primarily) is apt to cause behaviour of a certain kind (*MTM*, 82). This, he suggests, is backed by physiological psychology, which gives an account of persons as objects that are 'continually acted upon by certain physical stimuli. These stimuli elicit from [some person] certain behaviour, that is to say, a certain physical response. In the causal chain between the stimulus and the response, falls the mind. The mind is that which causally mediates our response to stimuli' (*MTM*, 78). As we saw above, this is one place where science aids conceptual analysis.

When it comes to behaviourism, he thinks one of its advantages is that it does *not* provide translations of mental statements in terms of behavioural statements (*MTM*, 67). Clearly, he is not thinking of positivist behaviourism because that view proposes such translations, whether analytic or synthetic. Armstrong is reacting to the sort of Rylean behaviourist who proposes a genuine theory of the nature of mind. It might be true that Ryle understood his position as non-metaphysical, that he took himself to be working on a cure for the disease of Cartesian thinking (Ryle 1949, 9). But Armstrong was always against the idea that 'philosophy does not issue in any theories at all about the nature of reality' (*MTM*, 14). So any form of behaviourism worth entertaining is a version that offers a theory about the nature of mind. Thus, behaviourism is one form of materialism. Given Armstrong's conception of philosophy (*MTM*, 85), the behaviourist provides a genuine account of mental phenomena in terms of a system of behavioural concepts. We must remember that Armstrong is honestly considering the prospect of adopting behaviourism (although he was never a behaviourist at any point, unlike his fellow Australian materialists). Behaviourism is not a foil he used in the service of his preferred theory.

In the end, Armstrong argues that 'genuine Rylean behaviourism' is not a candidate theory because it does not fit the correct analysis of a mental state. Ordinary experience tells us that mental states are causes of behaviour, something that the behaviourist denies. Conceptually speaking, the behaviourist only gives us part of the story of our concept of mind, namely, the part that says behavioural outputs are intimately connected with mental states (*MTM*, 68). The real problem for behaviourists is that they do not conceive of dispositional properties as events or causes, because for them a disposition is not a state had by some object. On Armstrong's realist theory of dispositions, dispositions are reduced to some actual categorical basis of the given object.[12] What would be a disposition and so a non-cause for Ryle is a cause for Armstrong. Sometimes Armstrong argues from ontic premises for this view, such as his appeal to truthmaking (Armstrong 1969, 23), but at other times he alludes to linguistic considerations. He writes: 'The Realist view gains some support from ordinary language, where we often seem to identify a disposition and its "categorical basis"' (*MTM*, 86), and that 'It is linguistically proper, . . . to say that brittleness *is* a certain sort of bonding of the molecules of the brittle object' (Armstrong 1969, 26, his italics). However, appeals to what is linguistically proper are out of place for someone like Armstrong.

---

[12] According to other realist theories of dispositions (such as D. H. Mellor's), natural properties just are dispositions without any need for a categorical base. For Mellor's reflections on how *MTM* shaped his views about dispositions, the mind, and beliefs and desires, see Mellor (2015).

Armstrong's stronger case for central-state materialism, I argue, rests on his use of methodological principles. One candidate view he must consider is Place's view/ Smart's initial identity theory. The difference between Place/Smart and Armstrong is that Armstrong gives a materialist reduction of *all* mental concepts, whereas Place and Smart adopt a materialist reduction of sensory concepts only. It is agreed, then, that *some* mental concepts receive a reductive analysis in terms of brain states. So we have the kind *inner mental state* in our ontology, which we are entitled to use in explaining facts about the mind. However, Smart's original motivation to restrict the number of inner mental states entails an incorrect application of Ockham's razor. The correct application seeks to limit the number of *kinds* of entities, not the number of *instances* of kinds. So, Place's and Smart's half-way behaviourist-identity theory is not better off, economically speaking (*MTM*, 81).[13]

The over-arching argument for central-state materialism is what I call a serviceability argument. Start with Armstrong's argument from the supremacy of physics for materialism. It is technically neutral between genuine Rylean behaviourism and central-state materialism. If Place and Smart are right, we must admit inner mental states. Armstrong's criticism that quantitative parsimony makes no difference to the economy of Place's and Smart's ontology puts central-state materialism forward as the more suitable theory of the two to systematically explain mental phenomena. In addition, any suitable candidate must satisfy certain desiderata. Think of these desiderata as facts that must be explained: every theory must explain (1) the fact that minds are individuated; (2) the fact that certain mental states (and not others) are unified; and (3) the fact that mental states are ontically dependent on substances, etc. Central-state materialism is more probable than its rivals because it explains the facts expressed in the desiderata using fewer primitives. He argues that competing theories do not explain all the facts that need to be explained. If they do, they do so with a greater number of primitives or basic principles. In short, central-state materialism is an economical and serviceable (fruitful) hypothesis, which is a reason to believe it is true. Armstrong employs this inference to the best explanation-style of argument when he argues for a factualist ontology (Armstrong 1997). His reductive materialism is really an early exercise in metaphysics. It shapes the philosophical method that he appeals to in later work.

---

[13] Armstrong is drawing a distinction between qualitative and quantitative parsimony (although he does not use these labels). His criticism presupposes that quantitative parsimony is not to be taken seriously. This distinction was popularized by Lewis, who also recognized 'no presumption whatever in favor of quantitative parsimony' (Lewis 1973, 87). The distinction dates back to Donald C. Williams. He called quantitative parsimony the 'gross tonnage conception of logical parsimony' and argued that it is 'no true logical economy, or is a vanishingly unimportant kind' (Williams 1934, 89). Armstrong studied Williams' *Principles of Empirical Realism* by 1967, so he may have found the distinction there. A favourable reference later on to Williams' discussion suggests that Armstrong was affected by it (Armstrong 2007, 99).

3. Lewis and Australian Materialism

Lewis was an undergraduate at Swarthmore College from 1957 to 1962. He enrolled in his first philosophy course, taught by Jerome Shaffer, in Fall, 1958. With a preliminary taste for philosophy and an awareness of his philosophic talent he joined his parents on a one-year visit to Oxford in 1959–60. Under the tutorship of Iris Murdoch he studied Hume, Kant, Leibniz, and Moore, and read contemporary works by Ryle, H. L. A. Hart, P. F. Strawson, Ayer, J. L. Austin, Wittgenstein, and Wisdom. His sojourn at Oxford had an immediate effect. In various undergraduate essays he endorsed logical behaviourism as found (according to some interpretations) in Ryle and Wittgenstein, e.g., 'Logical Behaviourism and the Problem of Other Minds' (David Lewis Papers, C1520, 'St. Catherine's Society, Oxford, Papers', Box B–000696 Folder 5, Princeton University Library).[14] He even stated in his applications for graduate school that Ryle's *Concept of Mind*, Wittgenstein's *Philosophical Investigations*, and Wisdom's *Philosophy and Psycho-Analysis* taught him the most thus far in his philosophical development (Beebee and Fisher 2020a, 6–10).

In early 1963, after his first semester at graduate school at Harvard University, he corresponded with Shaffer about the identity theory. It is at this point that Lewis encountered Smart's identity theory, not when Lewis and Smart first met at Harvard in Fall, 1963. Shaffer posed a number of problems for Smart's identity theory. One problem is that brain processes are locatable whereas mental states are not. If the identification of mental states with brain processes requires spatio-temporal coincidence, mental states are not identical (contingently) with brain processes (Shaffer 1961, 815–16). In a letter to Shaffer dated 12 March 1963, Lewis argues that Shaffer's objection shows us that the identity theory should not be formulated as asserting an identity between mental states and brain processes but rather between mental *states* and brain *states*. Each state is of something that has location. This something, a person, is the bearer of both mental and physical states (Beebee and Fisher 2020b, 3–4). (For Lewis, the word 'state' is a neutral term that covers universals, kinds of events, general attributes, types, not particulars, particular events, tokens, etc.)

---

[14] Hereafter, references to material in the Lewis Papers are shortened to folder name, box, and folder number.

It is clear from his correspondence with Shaffer that Lewis was attracted to the identity theory and less interested in saving behaviourism. Lewis brings this initial attraction and his materialist tendencies to Smart's seminar at Harvard in the Fall semester of 1963. After this semester Smart and Lewis discuss the identity theory in correspondence. Lewis modified Smart's topic-neutral account of sensation-reports as early as October 1964 (Letter from J. J. C. Smart to David Lewis, 8 October 1964, 'Smart, J. J. C.', Box B–000675 Folder 1). His proposal, as he says in his 1966 article, is that the topic-neutral account 'must specify typical causes or effects of the experience, not mere accompaniments' (Lewis 1966, 20). When I say 'the thing going on in me is like what goes on when ___', the blank must describe some cause or effect; whereas, for Smart, causation enters into the picture in order to determine which brain processes are mental states and which are not. This is not the same thing as analytically defining experience-ascriptions in terms of causal roles. Indeed, Smart shied away from claiming that the topic-neutral account is a *translation* of sensation statements in terms of physical statements (Smart 1967, 91–2).

The motivation for the identity theory rested largely on Ockham's razor. For many, it is too crude a device to wield in philosophy of mind. It cannot decide which theory wins. It can only screen off unworthy candidates. At the time, Michael C. Bradley criticized Smart's use of Ockham's razor, arguing that it is 'powerless in the kind of philosophical inquiry that Smart is engaging in' (Bradley 1964, 282). Lewis sought to undermine the need to appeal to parsimony by showing that the identity theory is implied by our best physiological theory. All we need to do is specify that some mental state M is that which plays some causal role C, allow science to discover that some brain state P plays role C, and conclude that mental state M is identical with brain state P.

Lewis's first premise is a generalization of Smart's topic-neutral analysis of sensation-reports. But Lewis remarks that the principle is behaviourist in origin; 'it inherits the behaviorist discovery that the (ostensibly) causal connections between an experience and its typical occasions and manifestations somehow contain a component of analytic necessity' (Lewis 1966, 21). Lewis's second premise is based on the working hypothesis that our best scientific theories are unified in such a way that physics and its laws entail and explain higher-order sciences and any laws of higher-order sciences. Here he leans heavily on (Oppenheim and Putnam 1958). It is interesting to note that later on, in reflecting on this paper, he remains more or less committed to the unity of science programme. He says to Philip Kitcher, on 2 January 1998, 'I don't disagree outright with theses that Oppenheim and Putnam assert, but some of the presuppositions in the background strike me as alien' (Beebee and Fisher 2020b, 189).[15]

---

[15] For recent criticism of this part of Lewis's argument, see Michael (2012).

While developing his view, Lewis reports to Smart that he is aware of the objection that the identity theory is false because if it is true, it is necessarily true, which contradicts their assertion that it is contingently true. Lewis says the objection rests on an unfair principle of individuation of states (or attributes)—namely, state F and G are identical iff necessarily F iff G. This sort of individuation principle might work for objects such as the evening star and morning star, but not for certain definite descriptions of states. To see this consider the property identity statement:

> The colour of the sky is identical with the property of being blue.

As Neil L. Wilson argued, if blue is the meaning of 'the property of being blue', it is the meaning of 'the colour of the sky'. If that is the case, this property identity statement is necessarily true. But it is contingent that the colour of the sky is the property of being blue. Therefore, blue is not the meaning of 'the colour of the sky' and as a result blue is not the meaning of 'the property of being blue'. Of course, we might say that both descriptions refer to the same property but the corresponding sentences do not mean the same thing (Wilson 1964, 53).

This example shows us that certain identity statements that involve definite descriptions of states are contingent. For Lewis, the statement:

> the state that is usually caused by stimulus S and that usually causes P-behaviour

is a definite description of some brain state, but we identify this brain state via a definite description of its contingent features. If the definite description is essential, states can be individuated in terms of necessary coextensiveness. If not, not. For example, the definite description 'being the state in which C-fibres firing is occurring' is an essential definite description of C-fibres firing (Beebee and Fisher 2020b, 8–11).

Lewis had come to realize that he needed a distinction between mental states and the havings of mental states. Mental states are identical with brain states but the havings of mental states are not. This was driven by the fact that 'the colour of the sky' contingently refers to the property of being blue but 'having the colour of the sky' necessarily refers to whatever colour the sky has (in every world that contains the sky). Lewis says of his term paper for Williams' metaphysics course (Spring, 1965), which became his 1966 article, that:

> The highly intensional stuff in the second half of the first part gives me a pain. I think it's unavoidable, however, if we're to avoid letting the I.T. become indistinguishable from the kind of epiphenomenalism that says that neural states are perfectly correlated with experiences. I do end up tolerating a purely logical sort of dualism, namely the dualism of experiences=neural states vs. havings of

> experiences. Notice that the same logical dualism turns up elsewhere in purely physical contexts; it's just one of the obnoxious features of attribute theory. When I first came upon the argument for it, however, I thought I'd have to abandon the I.T. At that point I just had the neural states vs. the havings of experiences (attributes expressed by experience-ascriptions) and hadn't thought to distinguish the experiences from the havings of them, as I later did. By the time I finished the first part of the paper I felt like turning nominalist for life!
>
> (Beebee and Fisher 2020b, 14–15)

After the 1966 article, there was a development of Lewis's theory that proved to be an important factor in shaping his philosophy. This was the realization that Smart's topic-neutral account of sensation-reports can be interpreted as a Ramsey sentence. In a letter to Smart on 16 May 1967, Lewis asks: 'Had you noticed that our topic-neutral translations of mental sentences are Ramsey sentences?' (Beebee and Fisher 2020b, 17). Folk psychology embodies claims such that 'I am in pain' is translated as 'I am in a specific state S_i such that there is a unique set or tuple of states $S_1 \ldots S_i$ such that these states satisfy the trivial claims of folk psychology.' (One difference between Lewis and Armstrong is that Armstrong appeals to physiological psychology to help support his causal analysis of the *general* concept of a mental state, whereas Lewis is happy to find the allusion to causal connections in the meanings of folk psychological terms. This is not to say that Armstrong was an empirical functionalist; Lewis and Armstrong were both analytical or common-sense functionalists.) This allowed Lewis in 'Psychophysical and Theoretical Identifications' to identify mental states with brain states using a 'general hypothesis about the meanings of theoretical terms: that they are definable functionally, by reference to causal roles' (Lewis 1972, 249), which he laid out in 'How to Define Theoretical Terms' (Lewis 1970).

Lewis's doctrine of theoretical terms derives from Ramsey's work on theoretical terms and the then commonplace distinction between theoretical and observational terms. For Lewis, the relevant distinction is between new or newly introduced terms by some theory T and old terms (not observational terms) of T. 'Pain' and other mental terms are newly introduced by specifying a certain role and associating it with the appropriate word, in this case, 'pain'. Pain is said to be whatever occupies that role. It just so happens that a certain brain state, say, C-fibres firing realizes a portion of folk psychology by filling the pain-role. The term 'pain' is a non-rigid designator that denotes the kind or universal pain (not token states). By contrast, the property of having pain is a rigid designator; the property of having pain non-contingently names the property in some world w belonging to whatever entities have pain in w.

If central-state materialism is true, 'pain' will turn out non-rigid. It is no objection to argue that central-state materialism is somehow wrong because 'pain' is in fact rigid. For there is no clear intuition independent of theory that supports this linguistic thesis. Such linguistic data does not sit pre-theoretically outside competing theories about the mind (see Lewis 1980, 218). Furthermore, Lewis thinks mental terms semantically resemble newly introduced terms, not that mental terms are theoretical *simpliciter*. Their semantic feature is such that it is as if they were newly introduced terms. Hence the empirical claim that in natural language the ordinary proper name 'pain' is rigid has no force. This issue, in short, is downstream from what ultimately determines the correct theory of mind.

Now, 'Psychophysical and Theoretical Identifications' was written before 'How to Define Theoretical Terms'.[16] This indicates that although his theory of mind is an instance of his doctrine of theoretical terms, Lewis arrived at the general doctrine by working through his theory of mind. This makes his philosophy of mind important for understanding the origins of his philosophical thought. For his doctrine of theoretical terms underpins his philosophical theorizing. He often begins his theorizing about some philosophical topic by claiming that the use of some term is unsettled and that it can be associated with many theoretical roles in our thought and philosophical theories. Something should earn its name by filling some specified role associated with the name. Something should not be given a name by saying that it should be so called, as Lewis once said jokingly you cannot have 'mighty biceps just by being called "Armstrong"' (Lewis 1983, 366). He tacitly operates with these methodological rules throughout his work—from ontology to value theory (for one relevant discussion, see Fisher 2018).

According to my interpretation, Lewis was originally a Rylean behaviourist who modified and generalized certain aspects of Smart's identity theory and applied Ramsey's method of defining theoretical terms to mental terms. Positivist behaviourism does little work in Lewis's theory. However, later developments of positivist behaviourism, found, for instance, in Carnap's work in the 1950s, also appealed in some way to Ramsey's method, although of course within a logical positivist framework and in accordance with their conception of a scientific theory. On this view, behaviourism remains an application of the positivist's Physicalism to psychology, but mental terms such as 'pain' are now theoretical terms, partially defined in terms of observational terms of a certain scientific theory (Carnap 1956, 69–75). Incidentally, it was this sort of positivist behaviourism that Putnam sought to reject (not its earlier manifestation from the 1930s), arguing that the meaning of 'pain' does not involve analytic connections between pain and pain-behaviour; 'pain' is not merely defined in this way; for one thing, it has a reporting use that is not implicitly defined by some theory (Putnam 1975, 449–50). Is Carnap's later behaviourism a source for and influence on Lewis? After all, Lewis's doctrine of theoretical terms is informed by Carnap's work on theoretical terms. If this were the case, it might be a connection between positivist behaviourism and Lewis's identity theory that should not be overlooked.[17]

---

[16] The former was read in March 1968, despite appearing in 1972 (and in Cheng 1975). Lewis writes to Smart on 20 April 1968: 'Cheng has not yet sent me the tape of my talk. I'd like to have that in the book [Cheng 1975] with as little change as possible; then write the stuff up as a paper called "How to Define Theoretical Terms" with some new material, and only a passing mention of mental states' (Beebee and Fisher 2020b, 22).

But I do not think there is a strong connection in the right place. First, as I explained, Lewis's general doctrine of theoretical terms develops out of his theory of mind, not vice versa. So, if the suggestion is that Carnap's influence on Lewis's doctrine of theoretical terms informs Lewis's theory of mind, it has the interpretation of the development of Lewis's theory of mind backwards. Second, Lewis had already signed on to Rylean behaviourism and from this position reasoned his way to the identity theory. So it is false that he began with some overt form of positivist behaviourism. Third, there are crucial differences between Lewis's theory of mind and Carnap's (later) view. For instance, it is part of Lewis's formulation of the Ramsey sentence in folk psychology that the functional definitions of mental terms are derived from common sense, whereas Carnap would hand the functional specification of mental terms over to empirical psychology (Crawford 2013, 638). Fourth, Lewis's first encounter with Ramsey comes not from Carnap but rather from Israel Scheffler in early 1964 in Scheffler's philosophy of science course. Certain elements of Lewis's interpretation actually stem from a reaction to Scheffler's reading of Ramsey. Since this has not been canvassed elsewhere, it is instructive to look at it in a bit more detail.

In that course Scheffler assigned his book *The Anatomy of Inquiry*, wherein he presents Ramsey's theory (Scheffler 1963, 203–22). Scheffler thinks Ramsey's method offers a way to eliminate theoretical terms in a fictional manner; he labels it 'eliminative fictionalism'. On this reading, Ramsey's method says that certain terms of a theory, say, a scientific theory, are meaningless or false. This is one traditional interpretation of Ramsey's method in the service of one form of behaviourism: the view that says there are no minds or mental events and thereby eliminates mental terms.

While this might have been one, perhaps common, understanding of Ramsey's method, it was not Lewis's.[18] Lewis draws a distinction between eliminating terms from the set of *primitive* terms in the theory and eliminating the ontic commitments found in those terms. In the case of mental vocabulary, Lewis Ramseyifies out mental terms, and as a result, mental terms are eliminated from the primitive vocabulary. But they remain available in the sense that we still have such expressions in our theory and they denote mental states. The expression 'that which fills the pain-role' denotes pain. Ramsey's method is meant to be a 'vindication' of theoretical terms (Lewis 1970, 427). Lewis did not think of Ramsey's method as supporting the view that eliminated terms are false or meaningless.

---

[17] It is beyond the scope of this chapter to explain how Lewis's doctrine of theoretical terms is informed by Carnap. But, briefly, what Lewis developed were certain aspects of the theoretical machinery of Carnap's work on theoretical terms. The question I am presently considering concerns the potential connection between positivist behaviourism (the version that says 'pain' is a theoretical term) and Lewis's identity theory.

[18] It was probably not Carnap's either. He admits criteria of signification for theoretical terms of psychology (Carnap 1956, 49–52).

Interestingly, Lewis's interpretation of Ramsey's method not only motivates his reductive materialism but also his rejection of eliminativist theories in general. An entity can deserve a name imperfectly so long as it is good enough to fill the corresponding role (and so deserves that name). To be sure, this depends on what the role-filler is, but for there to be some sort of flexibility in role-filling and an imperfection in name-deserving, there has to be a rejection of what Lewis calls 'semantic perfectionism'.[19] Thus, the choice is between rejecting semantic perfectionism and accepting eliminativism. Lewis chooses the first option (i.e., rejects semantic perfectionism). Language betrays a harmless and uncontroversial semantic indeterminacy and our definitions should reflect this vagueness and indeterminacy. In the case of mind, brain states occupy the role of mental states well enough such that they deserve the name of the mental state. *If* brain states are not good enough in filling the pain-role and so not good enough to deserve the name 'pain', Lewis would accept that, *strictly speaking*, there are no pains. Loosely speaking, there are pains if in the loose sense brain states are good enough to deserve the name 'pain'. Semantic indecision and ontology thus meet halfway.

In sum, from Lewis's perspective, he is reacting in part to Scheffler's treatment of Ramsey. Carnap's work on theoretical terms does inform Lewis's doctrine of theoretical terms (especially its theoretical machinery), but it does not assist Lewis on his path from behaviourism to the identity theory. Lewis's starting point is his own Rylean behaviourism. Hence Lewis is more appropriately located alongside Place, Smart, and Armstrong in the Australian materialist tradition. Finally, like Armstrong, Lewis's philosophy of mind is the initial project from which he extracts general tools and methodological rules that he applies elsewhere in his philosophy.

4. Conclusion

Australian materialism is a distinctive movement in the history of analytic philosophy of mind. It is separate from Feigl's identity theory and distinct from the evolution of positivism. It has its origins in a certain interpretation of Rylean behaviourism, namely the interpretation that behaviourism offers a genuine account of the nature of mind in terms of behaviour and behavioural concepts. There are separate lines of influence and distinct origins of the identity theory in analytic philosophy. Also, there are different explanations of the transition from behaviourism to the identity theory. Australian materialists never understood the issue as a clash between pseudo-problem and genuine problem (nor did they set it up in these terms). From the outset they considered the mind-body problem as genuine and interpreted logical behaviourism as offering a genuine view of the nature of mind. From this position, as I have explained, they travelled some distance to the identity theory. The more significant move was drawing a distinction between conceptual and factual truths and realizing that what was worth saving from logical behaviourism was its conceptual truths about mental states and its contributions to the conceptual part of folk psychology. Australian materialism is also significant because it shaped Armstrong's and Lewis's philosophy and philosophical method. Armstrong's metaphysics and metaphysical method date back to his arguments in *MTM*. Lewis's philosophical method and his systematic use of his Ramseyan method of defining theoretical terms stem from his identity theory. Lastly, since Australian materialism was one of the first metaphysical theses in the mid-twentieth century, it should be seen as an early contribution to the revival of metaphysics.

---

[19] In a letter to Christopher Jones, 10 November 1992; 'J', Box B–000667 Folder 12 (which has not been recently published).

Where should we locate Australian materialism in the history of philosophy, broadly speaking? Armstrong, along with Brian Medlin (1967, 95), dubbed the view 'central-state materialism' for a reason. To be sure, it was inspired by Feigl's label 'central-state theory', but it was intended to capture the fact that the view is a variety of *materialism*—a view that was scoffed at in the early twentieth century (e.g., Fullerton 1904, 247–57) and ridiculed by logical positivists for being unscientific, metaphysical speculation. Australian materialism is not to be associated with so-called 'identity' theories or 'mind-body monistic' world-views, some of which, in Armstrong's eyes, should be called double-aspect accounts of the mind. I submit that we should place Australian materialism within the larger materialist tradition that dates back to Hobbes and further to Lucretius, broadly speaking, and consider it as the latest episode in the long history of materialism cut using tools sharpened by the kind of analytic philosopher who proposes a genuine and systematic theory of the nature of things.[20]

References

Alexander, Samuel (1920) *Space, Time, and Deity: The Gifford Lectures at Glasgow 1916–1918*, 2 vols, London: Macmillan.

Anderson, John (2005) *Space-Time and the Proposition: The 1944 Lectures on Samuel Alexander's Space, Time, and Deity*, edited by Mark Weblin, Sydney: Sydney University Press.

---

[20] Thanks to the audience at the *A Materialist Theory of the Mind: 50 Years On* conference, University of Sydney, 10 August 2018. I also thank Peter Anstey, Keith Campbell, Sean Crawford, John Heil, and the late Hugh Mellor for comments and discussion. I am grateful to the late Steffi Lewis for permission to publish excerpts of letters from David Lewis, courtesy of Princeton University Library. This research was supported by the AHRC through its funding of *The Age of Metaphysical Revolution: David Lewis and His Place in the History of Analytic Philosophy* [grant number AH/N004000/1].

Anderson, John (2007) *Space, Time and the Categories: Lectures on Metaphysics 1949–50*, edited by Creagh Cole, Sydney: Sydney University Press.

Armstrong, D. M. (1954) 'Berkeley's Puzzle about the Water That Seems Both Hot and Cold', *Analysis*, 15(2): 44–6.

Armstrong, D. M. (1955) 'Illusions of Sense', *Australasian Journal of Philosophy*, 33(2): 88–106.

Armstrong, D. M. (1960) *Berkeley's Theory of Vision: A Critical Examination of Bishop Berkeley's Essay Towards a New Theory of Vision*, Melbourne: Melbourne University Press.

Armstrong, D. M. (1961) *Perception and the Physical World*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1962) *Bodily Sensations*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1966) 'The Nature of Mind', *Arts* (Proceedings of the Sydney University Arts Association), 3: 37–48.

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1969) 'Dispositions Are Causes', *Analysis*, 30(1): 23–6.

Armstrong, D. M. (1978) *Universals and Scientific Realism*, 2 vols, Cambridge: Cambridge University Press.

Armstrong, D. M. (1984) 'Self-Profile', in *D. M. Armstrong, Profiles*, vol. 4, edited by Radu J. Bogdan, Dordrecht: D. Reidel, 3–51.

Armstrong, D. M. (1989) 'C. B. Martin, Counterfactuals, Causality, and Conditionals', in *Cause, Mind, and Reality: Essays Honoring C.B. Martin*, edited by John Heil, Dordrecht: Kluwer, 7–15.

Armstrong, D. M. (1993a) 'Preface', in *A Materialist Theory of the Mind*, 2nd edition, London: Routledge, xiii–xxiii.

Armstrong, D. M. (1993b) 'Reply to Smart', in *Ontology, Causality and Mind*, edited by John Bacon, Keith Campbell, and Lloyd Reinhardt, Cambridge: Cambridge University Press, 169–74.

Armstrong, D. M. (1997) *A World of States of Affairs*, Cambridge: Cambridge University Press.

Armstrong, D. M. (2007) 'Truthmakers for Negative Truths and for Truths of Mere Possibility', in *Metaphysics and Truthmakers*, edited by J.-M. Monnoyer, Heusenstamm, Germany: Ontos Verlag, 99–104.

Ayer, A. J. (1936) *Language, Truth, and Logic*, London: Penguin Books.

Bacon, John, Keith Campbell, and Lloyd Reinhardt, eds (1993) 'Preface', in *Ontology, Causality and Mind*, edited by John Bacon, Keith Campbell, and Lloyd Reinhardt, Cambridge: Cambridge University Press, vii–xi.

Baker, A. J. (1986) *Australian Realism: The Systematic Philosophy of John Anderson*, Cambridge: Cambridge University Press.

Beebee, Helen and A. R. J. Fisher, eds (2020a) *Philosophical Letters of David K. Lewis: Vol. 1: Causation, Modality, Ontology*, Oxford: Oxford University Press.

Beebee, Helen and A. R. J. Fisher, eds (2020b) *Philosophical Letters of David K. Lewis: Vol. 2: Mind, Language, Epistemology*, Oxford: Oxford University Press.

Bradley, M. C. (1964) 'Critical Notice of *Philosophy and Scientific Realism*', *Australasian Journal of Philosophy*, 42(2): 262–83.

Carnap, Rudolf (1956) 'The Methodological Character of Theoretical Concepts', *Minnesota Studies in the Philosophy of Science*, 1(1): 38–76.

Cheng, Chung-ying, ed. (1975) *Philosophical Aspects of the Mind-Body Problem*, Honolulu: University Press of Hawaii.

Crawford, Sean (2013) 'The Myth of Logical Behaviourism and the Origins of the Identity Theory', in *The Oxford Handbook of the History of Analytic Philosophy*, edited by Michael Beaney, Oxford: Oxford University Press, 621–55.

Feigl, Herbert (1958) 'The "Mental" and the "Physical"', *Minnesota Studies in the Philosophy of Science* 2: 370–497.

Feigl, Herbert (1975) 'Some Crucial Issues of Mind-Body Monism', in *Philosophical Aspects of the Mind-Body Problem*, edited by Chung-ying Cheng, Honolulu: University Press of Hawaii, 20–34.

Fisher, A. R. J. (2015) 'Samuel Alexander's Theory of Categories', *The Monist*, 98(3): 246–67.

Fisher, A. R. J. (2018) 'On Lewis Against Magic: A Study of Method in Metaphysics', *Synthese*, 195(5): 2335–53.

Franklin, James (2003) *Corrupting the Youth: A History of Philosophy in Australia*, Paddington, NSW: Macleay Press.

Fullerton, G. S. (1904) *A System of Metaphysics*, New York: Macmillan.

Grave, S. A. (1976) *Philosophy in Australia since 1958*, Sydney: Sydney University Press.

Grave, S. A. (1984) *A History of Philosophy in Australia*, St Lucia: University of Queensland Press.

Heidelberger, Michael (2001) 'The Mind-body Problem in the Origin of Logical Empiricism: Herbert Feigl and Psychophysical Parallelism', in *Logical Empiricism: Historical and Contemporary Perspectives*, edited by Paolo Parrini, Wesley C. Salmon, and Merrilee H. Salmon, Pittsburgh, PA: Pittsburgh University Press, 233–62.

Kim, Jaegwon (1998) *Mind in a Physical World*, Cambridge, MA: MIT Press.

Lewis, David (1966) 'An Argument for the Identity Theory', *Journal of Philosophy*, 63(1): 17–25.

Lewis, David (1970) 'How to Define Theoretical Terms', *Journal of Philosophy*, 67(13): 427–46.

Lewis, David (1972) 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy*, 50(3): 249–58.

Lewis, David (1973) *Counterfactuals*, Cambridge, MA: Harvard University Press.

Lewis, David (1980) 'Mad Pain and Martian Pain', in *Readings in Philosophy of Psychology*, edited by Ned Block, Cambridge, MA: Harvard University Press, 216–32.

Lewis, David (1983) 'New Work for a Theory of Universals', *Australasian Journal of Philosophy*, 61(4): 343–77.

Mandelbaum, Maurice (1955) 'Societal Facts', *British Journal of Sociology*, 6(4): 305–17.

Martin, C. B. (1959) *Religious Belief*, Ithaca, NY: Cornell University Press.

Martin, C. B. and Max Deutscher (1966) 'Remembering', *The Philosophical Review*, 75(2): 161–96.

McLaughlin, Brian P. and Ronald Planer (2014) 'The Contributions of U. T. Place, H. Feigl and J. J. C. Smart to the Identity Theory of Consciousness', in *Philosophy of Mind: The Key Thinkers*, edited by Andrew Bailey, London: Bloomsbury, 103–28.

Medlin, Brian (1967) 'Ryle and the Mechanical Hypothesis', in *The Identity Theory of Mind*, edited by C. F. Presley, St Lucia: University of Queensland Press, 94–150.

Mellor, D. H. (2015) 'From Translations to Truthmakers', in *Metaphysics and Scientific Realism: Essays in Honour of David Malet Armstrong*, edited by F. F. Calemi, Berlin: De Gruyter, 219–31.

Michael, Michaelis (2012) 'Problems with Lewis' Argument for the Identity Theory', *Ratio*, 26(1): 51–61.

Murdoch, Iris (1956) 'Vision and Choice in Morality', *Aristotelian Society Supplementary Volume*, 30(1): 32–58.

O'Connor, D. J. (1963) 'Review of *Berkeley's Theory of Vision*', *Journal of Philosophy*, 60(16): 472–3.

Oppenheim, Paul and Hilary Putnam (1958) 'Unity of Science as a Working Hypothesis', *Minnesota Studies in the Philosophy of Science*, 2: 3–36.

Pepper, Stephen C. (1975) 'A Split in the Identity Theory', in *Philosophical Aspects of the Mind-Body Problem*, edited by Chung-ying Cheng, Honolulu: University Press of Hawaii, 35–42.

Place, U. T. (1954) 'The Concept of Heed', *British Journal of Psychology*, 45(4): 243–55.

Place, U. T. (1956) 'Is Consciousness a Brain Process?', *British Journal of Psychology*, 47(1): 44–50.

Presley, C. F. (1967) 'Introduction', in *The Identity Theory of Mind*, edited by C. F. Presley, St Lucia: University of Queensland Press, xi–xix.

Price, H. H. (1932) *Perception*, London: Methuen and Co.

Price, H. H. (1940) *Hume's Theory of the External World*, Oxford: Oxford University Press.

Price, H. H. (1953) *Thinking and Experience*, Cambridge, MA: Harvard University Press.

Putnam, Hilary (1975) 'Logical Positivism and the Philosophy of Mind', in *Mind, Language and Reality*, edited by Hilary Putnam, Cambridge: Cambridge University Press, 441–51.

Ryle, Gilbert (1949) *The Concept of Mind*, London: Hutchinson.

Ryle, Gilbert (1951) 'Feelings', *Philosophical Quarterly*, 1(3): 193–205.

Scheffler, Israel (1963) *The Anatomy of Inquiry: Philosophical Studies in the Theory of Science*, London: Routledge.

Searle, John (2004) *Mind: A Brief Introduction*, Oxford: Oxford University Press.

Shaffer, Jerome A. (1961) 'Could Mental States Be Brain Processes?', *Journal of Philosophy*, 58(26): 813–22.

Smart, J. J. C. (1955) 'Critical Notice of *From a Logical Point of View*', *Australasian Journal of Philosophy*, 33(1): 45–56.

Smart, J. J. C. (1957) 'Plausible Reasoning in Philosophy', *Mind*, 66: 75–8.

Smart, J. J. C. (1959) 'Sensations and Brain Processes', *The Philosophical Review*, 68(2): 141–56.

Smart, J. J. C. (1967) 'Comments on the Papers', in *The Identity Theory of Mind*, edited by C. F. Presley, St Lucia: University of Queensland Press, 84–93.

Stubenberg, Leopold (1997) 'Austria vs. Australia: Two Versions of the Identity Theory', in *Austrian Philosophy, Past and Present: Essays in Honour of Rudolf Haller*, edited by Keith Lehrer and Johann Christian Marek, Dordrecht: Kluwer, 125–46.

Tanney, Julia (2009) 'Rethinking Ryle: A Critical Discussion of *The Concept of Mind*', in *The Concept of Mind: 60th Anniversary Edition*, edited by Julia Tanney, London: Routledge, ix–lvii.

Williams, Donald C. (1934) 'The Inductive Argument for Subjectivism', *The Monist*, 44(1): 80–107.

Wilson, N. L. (1964) 'The Trouble with Meanings', *Dialogue*, 3(1): 52–64.

Wisdom, John (1953) *Philosophy and Psycho-Analysis*, Oxford: Basil Blackwell.

Wolfe, Charles T. (2015) *Materialism: A Historico-Philosophical Introduction*, Cham: Springer.

4

Causation, Perception and Dispositions

Towards *A Materialist Theory of the Mind*

Sharon Ford

Introduction

One of the innovative features of Armstrong's insight in *A Materialist Theory of the Mind* (hereafter *MTM*) is to question what had become, since Gilbert Ryle's *The Concept of Mind*, a traditional way of viewing dispositions. Since Ryle, dispositions had been regarded as counterfactual explanations for what would or could happen when particular circumstances are encountered by a thing, including by things with minds. In response to Ryle, Armstrong attempted to provide a realist theory of dispositions (*MTM*, 85–8), an account of how the counterfactual claims attributed to dispositions could be warranted by reducing dispositions to the physical features or properties of the microstructure of their possessors. In this chapter, I explore the background influences that led to this innovative leap, which was, in turn, a formative influence in twentieth-century philosophy of dispositions and powers.

MTM was published relatively early in Armstrong's academic career, and hence there is comparatively little published work to draw on for exploring the development of his thought up to the time of writing and publication. This chapter uses the available published work, together with a range of Armstrong's unpublished manuscripts and communications, to fill out that story. Section 1 begins with work Armstrong undertook as an undergraduate student at Sydney University. One cluster of key concepts that he continues to develop through the entirety of his philosophic career—questions about causation, universals, and particulars, the uniformity of nature, universal and necessary connection, kinds and instances, and particulars and powers—exists in embryonic form in the earliest stages of his philosophic thought. Section 2 describes a second cluster of key concepts that arise with Armstrong's move to Birkbeck College, London, to take up an academic position. Here, the thematic focus of his writing changes to defend and develop what he calls the Naïve or Direct Theory of Perception (Armstrong 1984, 10). Section 3 then highlights the role of dispositions as a catalyst in the convergence of

Sharon Ford, *Causation, Perception, and Dispositions: Towards* A Materialist Theory of the Mind In: *Armstrong's Materialist Theory of Mind.* Edited by: Peter R. Anstey and David Braddon-Mitchell, Oxford University Press. © Sharon Ford 2022.
DOI: 10.1093/oso/9780192843722.003.0005

these two diverse thematic areas, causation and perception, to produce an account of mind that attempts to ground the mental in the physical.

1. Early Work on Causation

1.1 Undergraduate Papers: Sydney, 1949

In his early writings, Armstrong's key focus on causation begins with an analysis of David Hume's Regularity Theory of Causation, as Hume outlines it in his *A Treatise of Human Nature* (Part iii, Section ii). Armstrong focuses on the key concept of necessary connection that Hume draws upon in this theory. Armstrong's analysis begins with two undergraduate manuscripts, 'Nature of Belief' (Armstrong 1949a) and 'Science as Superstition' (Armstrong 1949b), which focus on the topic of uniformities, these being for Armstrong, central to science, noting that 'science, then, has as its object the discovery of "uniformities"' (Armstrong 1949b, 2).[1] These manuscripts unpack Hume's if-then conditional concerning causal connection; namely that *if* one supposes uniformity is tied to necessary connection, *then* you cannot by this means get from the particular to the universal. While Armstrong agrees with this conditional statement, his argument focuses on the claim that Hume's replacement of 'universal connection' (or 'uniformity') with 'necessary connection' is illegitimate. He further argues that it is this mapping, rather than the concept of causation itself, that is the source of the problem that Hume identifies.

The arguments of these two manuscripts attempt to answer Hume by shoring up an account of the uniformities in nature with some kind of ontological ground for universal connectedness. They use spatio-temporal connections as a key conceptual tool by replacing Hume's 'necessary connection' with the use of space and time as the grounding of uniformity. In 'Science as Superstition' Armstrong writes, 'The only ground of connection of kinds is space and time . . . in any particular collection we shall have only spatial and temporal relations between kinds and it is impossible to see how these relations could yield the doctrine of necessary connection' (Armstrong 1949b, 6). And later in the same manuscript, 'it is not clear how we could know a thing as having certain relations without knowing the related terms. But we do know all things as spatio-temporal and as kinds' (Armstrong 1949b, 13).

For Armstrong, uniformity requires, in addition to spatial and temporal relations, the further concept that there are 'sorts' and 'kinds' of particulars, and that these particulars possess qualities which are themselves instances of more general sorts and kinds (a forerunner to his later discussions on properties and types):

---

[1] Citations from unpublished manuscripts follow the original pagination.

> It is important to emphasize that all connections of particular things whether universal or particular (in the sense of 'some A are B') depends on the fact that the particular things are Sorts or kinds of things. IF we do not admit this then we have no basis at all for collecting together particulars and prefixing them with 'All' or 'Same'—these expressions would be meaningless and, furthermore, we could have only spatio-temporal connections between groups of particulars—only if things are INSTANCES of kinds will any other connections exist.
>
> (Armstrong 1949a, 4, Armstrong's capitalization)

Altogether, key concepts found throughout these undergraduate papers include necessary connection, universal connection, spatio-temporal relations, and particulars as kinds. Of special importance are the steps that Armstrong makes in replacing necessary connection with universal connection, and then grounding this universal connection in spatio-temporal relations.

1.2 Honours Thesis, Sydney, 1950

Armstrong's 1949a critique of Hume is extended in his Honours Thesis, which focuses on Kant's 'Transcendental Analytic' as it appears in the *Critique of Pure Reason*. He focusses, in particular, on the climax of the Transcendental Analytic, "Analogies of Experience" (Armstrong 1950, 2). Armstrong's Honours Thesis presents a continuation of his undergraduate focus, that while we cannot experience necessity, we do not need to rely upon the concept of necessity to shore up causation. Universal relations can be observed in the instances, and can be so perceived precisely because they are instances of a type or sort of connection (Armstrong 1950, 11–12). Armstrong's focus is on claims that causal connection is an 'unobservable relation', whether Humean or Kantian versions of the story are told, just so long as the idea of causal connection relies upon the concept of it being a necessary connection.

To get to his conclusion, Armstrong describes Hume's argument using the following schema (p. 12):

(i) The invariable sequence (A ➔ B)
(ii) Particular sequences where individual As are succeeded by individual Bs (A1 ➔ B1).

Hume's rejection of the invariable sequence (i) leaves only the particular sequence, in which necessity is unobservable. Kant's Copernican Revolution replaces the idea that 'all our knowledge must conform to objects', with the reverse concept, 'objects must conform to our knowledge' (Kant, Bxvi, as cited in Armstrong 1950, 14). It therefore builds an intrinsic connection between the invariable and particular sequence, one that affects the way we see it. Armstrong argues in response that if we want to find experience-based necessity, we would now have to observe a universal conjunction between the particular and the invariant relation to know about this intrinsicality, and we are faced with the same line of argument as outlined by Hume. In short, so long as a necessary connection is postulated, we cannot provide experiential justification for it, regardless of whether we see the invariable and particular sequences as connected relationally or intrinsically.

1.3 Between Honours (Sydney) and Bachelor of Philosophy (Oxford)

1.3.1 Manuscript: 'The Meaning of Free Thought'

In 1951, between his Honours year and heading off to Oxford to embark on a Bachelor of Philosophy degree, Armstrong was a member of the Sydney Free Thought Society, run by Professor John Anderson. In the manuscript, 'The Meaning of Free Thought' (1951), Armstrong attempts to set a 'positive account' of freethought [*sic*] (Armstrong 1951, 2) for the Society. His stance on Naturalism is defended in this manuscript, with a disavowal of superstition, his noting that, '[i]n general superstition tries to get away from a simple naturalist or spatio-temporal view of things' (Armstrong 1951, 3). All existents, including the mental, are within physical space and time, and there is a further emphasis on the theme that spatio-temporal relations underpin the uniformities in nature. Note that in a few short years, in his 1953 manuscript 'Identity of Indiscernibles', Armstrong will take this idea much further, in a thesis which puts forward spatio-temporal relations as the criterion for individuation (Armstrong 1953a).

1.3.2 Study Group on Space and Time

In 1952, Armstrong was also a member of a study group on space and time, held in Sydney. Along with John Mackie, John Maze, David Stove, Richard Taylor, J. B. Thornton, and R. Walters, he presented seminar papers and commentaries as part of that group (Armstrong 1952a; 1952b; 1952c; 1952d; 1952e). In several papers, Armstrong supports an absolutist position on space and time. This position changes over time, however, with the publication of his much more refined and detailed 1963 paper, 'Absolute and Relative Motion' (Armstrong 1963a). In this paper, the absolutist position is no longer seen by Armstrong as defensible, and he instead points to complexities in the relativist view that need addressing. Also, in 1952, the idea of qualities versus relations features as a prominent part of the discussion, and is a forerunner of later discussions on categorical properties and powers. Another theme that features in manuscripts at this time is the question of continuity, inclusive of issues around points versus infinite divisibility.

1.4 Bachelor of Philosophy, Oxford, 1952–54

Moving from Sydney to England, Armstrong enrolled in the two-year Bachelor of Philosophy degree at Oxford University, from 1952 to 1954. He describes a change in philosophical emphasis in the academic environment, from empirical metaphysics at Sydney University to linguistic, semantic, and conceptual analysis at Oxford University (Armstrong 1984, 9–11). Armstrong notes that the Oxford philosophy was unfamiliar (Armstrong 1984, 9), and that it was a struggle and shock, at first, for the primary philosophical consideration to consist in the use of words (Armstrong 1984, 10). Given this, it is not surprising to encounter themes of his early Sydney experience carrying over into the manuscripts produced in 1953 and 1954, during his Bachelor of Philosophy studies. Through much of this work, the thematic content focused, in addition to space and time, upon properties and kinds. In what follows, I unpack some key manuscripts that formed part of Armstrong's studies in his Bachelor of Philosophy degree.

1.4.1 Manuscript: 'Is Hume's Analysis of Necessary Connection in Terms of a Determination of the Mind Circular?'

A key writing on causation is a manuscript entitled, 'Is Hume's analysis of necessary connection in terms of a determination of the mind circular?'. In this manuscript, Armstrong revisits his earlier rejection of the concept of necessary connection (Armstrong 1953b). The argument begins with Hume's division between 'relations of ideas' and 'matters of fact' (Armstrong 1953b, 2), and brings in a discussion of language and propositions. Armstrong now also begins to further develop his thoughts on the indirect observation of universal connection. He does so via observations of the patterns emerging through repetition and kinds, and the idea that causal sequence involves a concept of kinds or types which is sufficient to yield a suitable view of causation (Armstrong 1953b, 4–5). Universals take on a central focus, along with a renewed emphasis on the observability of universal connection. Two other key components of Armstrong's later view find their genesis in this paper. The first concerns a distinction between 'universal connection' and 'universal connection of universals' (Armstrong 1953b, 8–9). This is only barely introduced in this paper. However, it gets picked up and incorporated later in Armstrong's discussion about universal types and differentiation of property levels. The second key component in Armstrong's discussion of Hume is the topic of inference-making and its justification, and the problem of induction that Hume points to in *An Enquiry Concerning Human Understanding* (Section 6; Armstrong 1953b, 7–10).

1.4.2 Manuscript: 'The Identity of Indiscernibles'

The manuscript noted earlier, 'The identity of indiscernibles' (Armstrong 1953a), is another important paper produced in Armstrong's Bachelor of Philosophy studies. This manuscript discusses Leibniz, and now begins to articulate Armstrong's view of particulars, types, and universals in much more detail than in previous writing. His focus is squarely on ontological rather than the epistemological concerns. His aim is to answer the question, 'What is it to be two different things?', and so his earlier ideas about particulars being kinds and sorts now moves into the broader theme of individuation, which is given in terms of non-repeatability, understood as numerical difference (Armstrong 1953a, 1a–2). Two things may be things of the same sort because they share key properties. Armstrong gives, as an example, 'Peter and Paul are different things, but they are things of the same sort, they are both men. In the language of traditional philosophy, they are different particulars instantiating the same universal' (Armstrong 1953a, 1). There follows throughout the paper a careful and detailed taxonomy of how properties stand to particulars, including discussions of individuation, parthood, persistence, and how individuals should be considered in terms of spatial and temporal considerations. Armstrong develops a central focus on how objects can be represented in purely spatio-temporal terms by outlining a suggestion for spatio-temporal 'contouring' of objects (although this is not developed in detail here). Armstrong's conclusion concerning individuality is to re-emphasize that 'individuals are individuals because they are space-times. Two individuals of the same sort (the two French Revolutions) are different individuals because they are different space-times' (Armstrong 1953a, 9, Armstrong's underlining). The implication here is that *individuality is something more than a thing's properties.* Properties involve different particulars being the same sort, but do not impact upon a thing's individuation. This is the concept which will underwrite Armstrong's later metaphysics of states of affairs, and the inclusion of haecceity into his later categoricalist view.

1.4.3 Manuscripts: 'Mill on Liberty' and 'The Nature of Pain'

Two other manuscripts written in 1953 include 'Mill on Liberty' (Armstrong 1953c) and 'The Nature of Pain' (Armstrong 1953d). In these manuscripts, Armstrong begins to talk about properties and particulars with respect to people and the nature of experience. Of these two, 'The Nature of Pain' is relevant to his development of laws of nature as it crosses over to a discussion on qualities. It is concerned with how to characterize qualities or characters, given the private nature of our experiences. Armstrong defines a quality (or a character) in terms of the fact of it being observable, and then seeks to show how this may be applied to what we now call qualia, for example, pain. His solution is to attempt to split up pain into two different parts: the qualities or characters (i.e., properties) of pain that occur in painful places, and the aspect of 'being in pain'. From here, he generalizes to all bodily sensations, and begins to draw a distinction between mental acts and bodily sensations, an account which will be developed further in his 1962 book, *Bodily Sensations*, and is a key consideration in the development of the Mind-Brain Identity Theory, presented in *MTM*.

1.4.4 Bachelor of Philosophy Thesis: 'The Nature of Complexity'

Armstrong's Bachelor of Philosophy thesis was presented in March of 1954, entitled, 'The Nature of Complexity'. The key theme is that whatever is complex is a combination of simple elements (Armstrong 1954a). In keeping with his earlier rejection of necessary connections, Armstrong again uses Hume's regularity theory of causation to argue against necessary connection, and to present an account of complexity that denies logical necessity between objects or events in the world (Armstrong 1954a, 4). He aligns himself with Hume's principle that there can be no necessary links between distinct entities (Armstrong 1954a, 20–1), and Armstrong's discussion around this principle makes it clear that at this stage of his career he holds to the mosaic picture of the world in terms of simple elements being ultimately distinct one from the other: 'And so any complex character must 'break up' into a combination of distinct characters. We must then finally reach real simple characters, for it would be nonsense to say that the characters that make up a complex character dissolve indefinitely in further distinctions' (Armstrong 1954a, 21). Note that this claim against characters dissolving indefinitely becomes a little fuzzier later in his writing, particularly when the topic of continuity is broached. But at this stage, his stance that 'simple characters are necessarily absolutely simple' (Armstrong 1954a, 24, Armstrong's underlining) is firm, and supplies the grounds for the later development of his combinatorialism. The assumption and defence of discreteness also occurs throughout the Bachelor of Philosophy thesis, which is an important concept that becomes central in later discussions on dispositions.

It is in this thesis that we really start to see the beginnings of what is to become a major part of Armstrong theoretical focus, that of the nature of dispositions, the role they play, and how they should be defined and described. And it is here that we begin to see a forging of considerations that will lead to his discussion of dispositions in *MTM*. At this stage, 1954, dispositions are not considered to be properties (or characters) because they are not 'actual' (Armstrong 1954a, 11–12). Yet, even here we see certain tensions beginning to arise. Dispositions are not yet accepted as properties proper; a particular quantity of petrol can be described as inflammable but does not have 'inflammability' as a property (Armstrong 1954a, 12–13).

This view of dispositions as non-actual character changes in the next phase of Armstrong's Realism concerning dispositions whereby, in response to the work of Gilbert Ryle, Armstrong identifies dispositions with the state of the categorical base of their bearer. This will be discussed in more detail in Section 3 of this chapter. I digress slightly by noting that later Armstrong's realist-reductionist view of dispositions further morphs into his better-known supervenient theory of dispositions, in which Armstrong moves from type-type metaphysical reduction to token-token reduction. In this later Contingent Identity Theory of dispositions, the identity of token dispositions and token microstructures occurs with respect to prevailing and contingent laws of nature.

Returning to Armstrong's (1954a) 'The Nature of Complexity', in this thesis we also find the beginning of a longer conversation on the topic of instantiation, along with some discussion on intrinsic versus relational properties (which at this stage, he calls 'characters'). With the influence of John Locke in the background concerning how ideas are signified (Armstrong 1954a, 12–13), Armstrong discusses intrinsic versus relational characters. He defines intrinsic characters of things or events as 'those characters which can be discovered without going beyond the thing or event in any way at all' (Armstrong 1954a, 13). In the case of an object in space, therefore, its intrinsic characters are 'those which can be discovered without going beyond the spatial boundary of the thing' (Armstrong 1954a, 13). In contrast, relational properties are 'characters of a particular which are not discoverable without going beyond the particular, and which do not characterize the particular intrinsically in any way' (Armstrong 1954a, 13). Armstrong sees the two, intrinsic and relational, as 'strictly exhaustive and exclusive' (Armstrong 1954a, 13).

Much of the argument in 'The Nature of Complexity' also trips over the topic of infinitudes, mathematical density (infinite divisibility) and spacetime points, to the conclusion that 'where we are dealing with intrinsic characters of particulars, complex characters must be combinations of absolutely simple characters' (Armstrong 1954a, 109). However, he argues that at least some spatio-temporal extensions and durations cannot be resolved into simples. Armstrong writes, 'It appears, then, that certain of the spatio-temporal characters of things (length, area, volume and duration), though complex, are not combinations of simple elements' (Armstrong 1954a, 76–7). This question involves the idea of continuity, and the view that there can be no gaps in time and space. The idea of a continuity in time and space is set against the idea of distinctness or discreteness, a tension that Armstrong explicitly recognizes he does not resolve in this paper (Armstrong 1954a, 86a, 109).

1.4.5 Manuscript: 'Infinite Divisibility of Time'

Further on the theme of space and time, 'Infinite Divisibility of Time' (Armstrong 1954b), discusses the idea of time (and space) as continuous or dense in the mathematical sense. The manuscript was completed after the submission of Armstrong's Bachelor of Philosophy thesis, dated June, 1954. It outlines three key issues that Armstrong argues against: (1) the argument from the senses; (2) the requirement for instances; and (3) the problem of infinity. The first argument comes from Berkeley and Hume, that all our knowledge is derived from our senses, that things are as they appear to be and appear as they are, in Armstrong's interpretation of this doctrine (Berkeley and Hume, as cited in Armstrong 1954b, 3–4). This means that we can know nothing but sensible duration, and that our senses do not permit the infinitely divisible. Armstrong's answer is to reject the premise of this argument. The second argument rejects infinite divisibility on the grounds that it prevents us from giving an intelligible account of instants (Armstrong 1954b, 4). The third argument involves key problems with infinity that also feature in problems concerning infinite divisibility. Armstrong's solution is to argue that while it might be hard to think about the world being actually infinite in certain ways, we cannot find any good reason for thinking otherwise. The difficulties represented in our theorizing about infinity are not such that we should dismiss it (Armstrong 1954b, 4–5). While all three issues are closely tied together in this manuscript, the first of these issues leads Armstrong directly to his first publication, 'Berkeley's Puzzle about the Water that Seems Both Hot and Cold' (Armstrong 1954c). This paper concentrates on providing an answer to Berkeley's puzzle concerning sense data. Focused around perception, it ties in to what will become for Armstrong a strong interest in perception and theory of mind. It is also through the avenue of his analysis of perception that Armstrong is drawn into the broader theory of dispositions, and this forms the starting place for Armstrong's realist account of dispositions.

2. Early Work in Perception

2.1 Birkbeck College, London University, 1954–56

In 1954, after completing his Bachelor of Philosophy at Oxford University, Armstrong took up a position at Birkbeck College, London. A key aim in his approach at this time is to defend a naïve realist theory of perception—later known as Armstrong's 'Direct Theory of Perception'—in contrast to not only the theories of historical figures, for example, Kant, but also prevailing sense-datum theories of his contemporaries, such as that of A. J. Ayer, who was also teaching at London University at that time.

As mentioned earlier, Armstrong's first publication of note came out in December of 1954, 'Berkeley's Puzzle about the Water that Seems Both Hot and Cold'. (This puzzle can be found in Berkeley's *Three Dialogues between Hylas and Philonous* (Berkeley 1949, 2, 178–9). In the case where one hand is hot and the other cold, the very same water may appear to this person as both cold and hot, respectively. In his short paper, Armstrong attempts to develop an answer to the puzzle, but one that avoids what he calls the 'desperate step' taken by the phenomenalist of the time, that of analysing the phenomena in terms of ideas or sense-data (Armstrong 1954c, 44). Instead, he concludes in line with John Anderson's solution, that the puzzle survives by dint of an incomplete description of the properties in question. The water is not hot and cold *per se*. These properties are not absolute properties of the water but are instead relative to different parts of the body.

In 1955, Armstrong published 'Illusions of Sense', the aim of the paper being to further defend a naïve realist theory of perception, this realist view involving the claim that 'different perceivers can, and very often do, perceive the very same thing' (Armstrong 1955, 88). In contrast, representative and phenomenalist theories deny this. In this paper, Armstrong does not provide a definition of a Representative Theory of Perception, but does note that these were, at the time, largely taken over by phenomenalism, a view that holds a physical object as a 'congeries or collection of sense-data, actual and possible' and that perceived physical objects are constructions out of this sense-data (Armstrong 1955).

2.2 Melbourne University, 1956

Armstrong' focus on perception continued through his transition from his position in London to his taking up a post at Melbourne University in 1956. Also important at this time is his publication on George Berkeley (Armstrong 1956a), which is a continuation of his earlier writing. In this publication, Armstrong attempts to explain how Berkeley arrives at his theory of sense impressions, his work on Berkeley culminating in his book, *Berkeley's Theory of Vision* (Armstrong 1960a). Between 1956 and 1958 he also produces a lecture series on F. H. Bradley's absolute idealism. The lectures focus on questions concerning whether circumstances that participate in allowing an object to be perceived are part of the object itself, and whether relations between objects and such circumstances are internal to an object. Bradley's work discusses important themes that are central to the development of Armstrong's later work, and in particular, questions about distinctness and discreteness, the status of objects and relations, the nature of truth and error, predication and substance. In 1956, Armstrong also produces a series of lectures on Immanuel Kant's *Critique of Pure Reason* (including sections from *Critique of Practical Reason*, and *Critique of the Power of Judgment* (Armstrong 1956b). Picking up the themes of his earlier writing and early manuscripts on Kant, the central themes of this lecture series are focused around the concepts of necessity and universality, on Kant's Copernican Revolution, his theory of perception, his views on space and time, and on Kant's categories (see Kant 1950; 1997; 2000).

Upon moving to Melbourne, Armstrong enrols in a doctorate, with A. C. Jackson as his supervisor, producing as its output the book *Perception and the Physical World* (Armstrong 1961b). This book forms the basis for further research in perception. The main thesis of the book draws together his arguments against the representative theory of perception and phenomenalism in favour of his direct realist position (a.k.a. Direct Perception) (Armstrong 1984, 17). Manuscripts written for inclusion in the book include 'The Location of Pain' (Armstrong 1958a), in which Armstrong continues to develop his stance against phenomenalism by reducing experiences of pain to physical properties. Another manuscript included in the book, 'The Nature of Perception', defends the stance that perception involves the acquiring of knowledge of particular facts about the physical world by means of the senses, and that perception is nothing more than this (Armstrong 1958b, 2). Other manuscripts include 'Causality' (Armstrong 1961a), 'Perception and the Physical World' (Armstrong 1960b), and 'Are the Primary Qualities Really Primary?' (Armstrong 1957–1958). Although the topic of focus throughout all the manuscripts around this time is perception, what has been achieved in 'Are the Primary Properties Really Primary?' is also a significant step into what will become a lifelong investigation and analysis of the relationship between things and their properties. Significantly, this manuscript supports Armstrong's view of direct perception by contending that qualities (or properties) are within things themselves, and not merely sense-data. Attempting to refine this account will later lead Armstrong into further exploring the nature of instantiation. But it is here, in this manuscript, that he appears to first seriously begin to forge a link between his theory of direct perception and a more general account of properties as inherent within things.

In his 1984 intellectual biography, Armstrong questions the degree of success of some of his arguments in his 1961 book. However, despite these shortcomings, he describes how he has nonetheless remained strong in his overall conviction against the idea of having to postulate sense-data in the explanation of perception. He persisted throughout his career in support of his theory of direct perception (Armstrong 1984, 15; 1997). Armstrong's rejection of sense-data theories is tied deeply to what appears to be a strengthening conviction that properties reside within the objects of the world, a theme revisited in a short monograph published in 1962, *Bodily Sensations*, in which he argues that bodily tactile sensations can be located in the physical world via the fact of our bodies being part of that world (Armstrong 1962). We are directly aware of tactile pressures on our body just as we are aware of colour or shape. It is at this point that he begins to merge the two main themes of his work: causation and perception, viewing direct perception as evidence that we directly perceive of causal force in bodily sensations.

3. Development of Armstrong's Reductive Analysis of Dispositions

Armstrong's description of dispositions in *MTM* involves the idea that dispositions are to be identified with the properties of the physical structures of the objects bearing them. Their engagement in the causal processes is made possible by way of this identification. The development of Armstrong's views on dispositions was part of his broader interest in the theory of properties which began around 1952 to 1954. His theory of properties is touched upon, for example, in his 1952 manuscript, 'Is Position in Space and Time Absolute or Relative?' (Armstrong 1952c). In this paper, Armstrong begins to separate out individuals from properties with the difference between the two based on the repeatability of properties. In 1953, we see further development of his concept of properties in terms of their instantiation in individuals and a disavowal of sequences as bare particulars (Armstrong 1953b, 5, 8). In this sense, 'bare particulars' are contrasted with particulars that are kinds or sorts of things, which for Armstrong means that they are 'instances of universals' (Armstrong 1953b, 8). Also, in 1953, in his manuscript 'The Nature of Pain' (Armstrong 1953d), and later in his book, *Bodily Sensations* (Armstrong 1962), Armstrong provides a definition of qualities (which he then calls characters). In 1954, in his Bachelor of Philosophy thesis, 'The Nature of Complexity', Armstrong begins to use the term 'character' interchangeably with 'properties' and for the first time, as far as his body of work demonstrates, he begins to think seriously about what may be considered dispositions versus what may be considered a quality (or character).

The context in which Armstrong was writing in the early 1950s was awash with discussion of dispositions. Gilbert Ryle had recently published *The Concept of Mind*, in which he argues against what he saw then as 'the official doctrine' in the philosophy of mind, Cartesian Dualism, Descartes' claim for the distinction between the mind and the body (see Ryle 1949, Chapter 1). Referring to the doctrine as the 'the dogma of the Ghost in the Machine', Ryle argues that performance of mental acts does not require a hidden layer of a mind which is separate from the performance, and causally antecedent to, the acts themselves. Ryle points out that many mental acts are dispositional in nature. For example, he writes: 'But in looking beyond the performance itself, we are not trying to pry into some hidden counterpart performance enacted on the supposed secret stage of the agent's inner life. We are considering his abilities and propensities of which this performance was an actualisation' (Ryle 1949, 45). This looks as though Ryle is seeing a performance or action as a manifestation of a set of underlying, and in some cases 'indefinitely heterogeneous' dispositions (Ryle 1949, 44).

Earlier verificationist philosophers prior to Ryle, such as Rudolph Carnap, had attempted to give an account of dispositions in a way that they could be observed and therefore verified. Carnap writes, 'Those predicates of the thing-language which are not observable, e.g. disposition terms, are reducible to observable predicates and hence confirmable. We have seen this in the example of the predicate "soluble"' (Carnap 1936, 466). The disposition could be described as a material conditional of the form:

$$(\mathrm{R}{:})\qquad (x)(t)[Q_1(x,\ t) \supset (Q_3(x) \equiv Q_2(x,\ t))]$$

which when described in words produces the reduction sentence: 'if any thing x is put into water at any time t, then, if x is soluble in water, x dissolves at the time t, and if x is not soluble in water, it does not' (Carnap 1936, 440–1). As Sungho Choi and Michael Fara note, such a description is silent concerning dispositions that are not in fact manifesting, which means that predicates pointing to the unobservable were still required in language (Choi and Fara 2006, 3). This was a problem for Carnap and other empiricists for it seemed to demonstrate that disposition ascriptions cannot, after all, be fully analysed by their being reduced to descriptions of the physical, observable world.

In response to this failed attempt to disavow language of dispositions, Ryle sought to soften the question of their legitimacy by tying their discussion to the language of counterfactuals: '[t]o possess a dispositional property is not to be in a particular state, or to undergo a particular change; it is to be bound or liable to be in a particular state, or to undergo a particular change, when a particular condition is realised' (Ryle 1949, 43). Ryle also extends this move to introduce dispositions as counterfactuals to human 'qualities of character' (e.g., being a smoker). Their counterfactual nature meant that disposition ascriptions no longer pointed to ontologically robust existences:

> It is now necessary to discuss briefly a general objection that is sometimes made to the whole programme of talking about capacities, tendencies, liabilities and pronesses. Potentialities, it is truistically said, are nothing actual. The world does not contain, over and above what exists and happens, some other things which are mere would-be things and could-be happenings. (Ryle 1949, 119)

The thinking was that if dispositions were to be considered only as 'would-be and could-be' ascriptions, their non-observability would no longer be of such relevant concern. Dispositions were to be thought of as 'inference tickets' (Ryle 1949, 117–25) for better understanding the repeated occurrence of observable patterns, which in the case of mind-stuff, included habitual mental acts and performances. In his earlier writing, arguably initially influenced by Ryle, Armstrong clearly rejects dispositions as actualized entities. That is, he does not consider dispositions to be ontologically robust, and is yet to formulate their grounding in categorical properties; this was to come much later, with his Inaugural Lecture, 'The Nature of Mind', given in 10 May 1965, and published in the Proceedings of the Sydney

University Arts Association (Armstrong 1966a). By contrast, in the mid-1950s, for example. as it appears in 'The Nature of Complexity', Armstrong is still very influenced by the Rylean account of dispositions.

> It is true that dispositional characters have a basis in the characters that particulars actually have (e.g. if we can truly say that this unignited sample of petrol is inflammable, then a great many samples of petrol must have actually ignited), but dispositions are not themselves actual characters of things.
>
> (Armstrong 1954a, 12)

Also, later in *MTM*, Armstrong recalls that prior to his discussions with Martin, he had thought of the concept of mind in line with Ryle's dispositional account, writing:

> Professor C. B. Martin, now my colleague at Sydney University, first made me aware of the central role played by the concept of *causality* in an account of the mental concepts. Previously I had assigned the central place to the notion of dispositions, conceived of as Ryle conceives them in *The Concept of Mind.* My change of view on this question was a turning point in my conception of philosophical psychology. (Armstrong 1968, xi)

Later, as Armstrong began to develop his realist account of dispositions, he opposed the Rylean view, indicating that Ryle seems to have earned the obligation to give an account of mental acts and performances that they are said to underlie (Ryle 1949, 87–8).

As made clear in the above quotation, Charlie Martin was a key influence in the development of Armstrong's view of dispositions. Martin took up a position at the University of Sydney in 1966 (Heil 2010, 260), working alongside Armstrong. This collegial relationship between Armstrong and Martin was a collaborative and productive one that continued throughout their lives. Later, in 1994, Martin published a paper called, 'Dispositions and Conditionals'. In a footnote (Martin 1994, 8), he mentions that he began this paper in 1957, took it up again in 1967, and finally, on the wings of friends' encouragement, published it in 1994. Armstrong is one person mentioned in his acknowledgements on this paper, and earlier unpublished versions of this paper appear among Armstrong's papers. Martin's (1957) manuscript comprises a discussion on the nature of dispositions, and the fact that these natures can, at least in theory, change. It also visits the relationship between dispositions and natural necessity.

Another example of Martin's influence on the development of Armstrong's early work on dispositions can be seen in 1958, the year in which Armstrong wrote 'The Location of Pain' (Armstrong 1958a). In its final form, this manuscript was included in his book, *Perception and the Physical World* (Armstrong 1961b). In the manuscript, Armstrong continues to develop his stance against phenomenalism by reducing experiences of pain to physical properties via the concept of belief, as follows, 'So now let us say that to have a pain in our hand is to believe or be inclined to believe on the basis of the sense of feeling that something physical is going on in our hand for which we have an immediate and interested dislike and which we wish would stop' (Armstrong 1958a, 24). Armstrong's conclusion in this manuscript is that pain is physical, but 'turns out to be a sub-species of the emotions of dislike or grief, dislike of or grief at certain physical happenings in our body, or more accurately, of certain putative physical happenings in our body' (Armstrong 1958a, 25). In a letter to Armstrong, Martin responds to this manuscript by pointing out that belief itself is largely dispositional, whereas a pain in the hand is not:

> 'Belief' itself is largely dispositional. We do not necessarily stop believing that something is the case just because we stop being aware of it. It may look as if 'on the basis of the sense of feeling' may take the weight off the non-dispositional side of 'pain'. But if 'belief' figures at all, won't it have to figure to some extent, and to that extent inappropriately, dispositionally? (Martin 1958)

Although Armstrong had studied with Ryle in his time at Oxford, dispositions do not show up as a major theme in Armstrong's work until his discussions with Martin around 1957 and this paper in 1958. The thrust of Martin's critique is that Armstrong's account of the experience of pain given in terms of the physical, turns out in fact to look very much like a phenomenalist account if belief is dispositional and dispositions are mere possibilities. Importantly, Martin's comments pushed Armstrong to a choice: either he gives up his physicalist account of mental events (such as the experience of pain) or else his dispositions need to be located somehow in the physical world. Martin's comments are therefore critical at this time in the development of Armstrong's realist account of dispositions.

A further influence in the development of Armstrong's realist account of dispositions includes his earlier naturalism. From the beginning of his written record, in manuscripts such as the 'Science as Superstition' (Armstrong 1949b), Armstrong's thought runs to the Aristotelian *doctrine of potential*, that unfulfilled possibilities cannot be considered as any 'addition of being'. In his manuscript 'The Nature of Perception' (Armstrong 1958b)—included as a chapter in *Perception and the Physical World* (Armstrong 1961b)—the argument against phenomenalism is carried over against A. J. Ayer's account (Ayer 1956, 112), interpreted by Armstrong as a reduction of unperceived objects to the 'merely hypothetical' or to the merely 'physically possible' (Armstrong 1961b, 55). This conceptual framework brings into focus questions concerning whether dispositions are themselves mere possibilities or something more. Though under-developed at that time, his view of dispositions now delivers them up as properties that can be acquired, and this is seen as a physical occurrent event (Armstrong 1961b, 121). If dispositions are properties rather than mere possibilities, then it would seem as though they should be treated as ontologically robust entities in the world. If Armstrong wanted to achieve an account of experiences (such as pain) compatible with his naturalism, then mental states, such as belief, needed to involve an account of dispositions in which these dispositions were treated as physical existents. When Armstrong writes, '[b]ut we can only attribute dispositions to things on the strength of certain non-dispositional characteristics' (Armstrong 1961b, 34), this gives us a brief glimpse of what was to later feature strongly in the arguments of *MTM* concerning the reduction of dispositions to the physical microstructure.

The topic of dispositions emerges more fully in Armstrong's writing at about the same time that he moves on from his core enterprise of defending direct realism to further development of his central-state materialism, as can be seen in another manuscript, 'Mind and Consciousness' (Armstrong 1961c), and in associated written notes, 'Mental Images' (Armstrong 1961d). The notes in 'Mental Images' seem to be an attempt to tease out the differences between sense impressions and mental images. According to a statement appended to 'Mind and Consciousness' (Armstrong 1961c, 1), *MTM* sprang from this important manuscript. In 'Mind and Consciousness', Armstrong begins to further clarify his mind-brain identity thesis: 'If we want a purely materialist theory of the mind, then let us put forward the hypothesis that the mind is the brain or is the central nervous system' (Armstrong 1961c, 1).

Throughout 1963 and 1964, we find Armstrong moving more strongly into this theme of mental images being physically robust, guaranteeing their actuality by identifying them with the physical. For example, in a letter written to J. T. Stevenson in December of 1963, Armstrong alludes to his Mind-Brain Identity Theory as follows:

> Incidentally, my materialism is not (at any rate orthodoxly) behaviouristic. In my view, the concept of a mental state is the concept of an 'inner' state of the person which is capable of producing behaviour. (The mind is a bit like the gene—mind stands to behaviour something as gene stands to actual inherited characteristics.) I then contingently identify this inner state with brain-states much as the gene is identified with the DNA molecule. (Armstrong 1963b, 2)[2]

The implications of this are brought to the foreground very strongly in Armstrong's Inaugural Lecture, 'The Nature of Mind' (Armstrong 1965). It is in this lecture that Armstrong's account of dispositions proper finds its origins.[3] Up to 1965, Armstrong had already begun to identify mind states with brain states, but the role and description of dispositions had been left dangling. It is with considerations of Ryle's work that Armstrong is now drawn into his first extended published discussion of dispositions. In 'The Nature of Mind' (Armstrong 1966a) Armstrong outlines what is missing in Ryle's account, namely an explanation for *why* things behave in the way they do, in the circumstances in which the behaviour takes place. Armstrong seeks to address this by providing his dispositions with a robust ontological presence. He does so by identifying them with the categorical base of the entities possessing them (*MTM*, 88–9; Armstrong 1980, 9; 1984, 22). By 'categorical microstructure', Armstrong means the everyday physical structure of an object. This identity relation between a disposition and the microstructure of an object appears as a two-way reduction whereby the disposition is *strictly identical* with the non-dispositional state of the object, that is, bonding of the molecules in their instantiating particulars (*MTM*, 88–9). Thus, when a disposition ascription is made, we are using a language shortcut to talk about the categorical basis. By 'non-dispositional' Armstrong is referring to something 'actual' as versus the non-existent and 'merely possible' account of dispositions.

---

[2] Armstrong attributes the gene analogy to Brian Medlin, see (*MTM*, 90).

Causation is brought in as a key concept in Armstrong's reductive account of dispositions (Armstrong 1966a; 1984, 22) By ascribing a causal role to the categorical base, talk of dispositions now becomes talk of how the categorical base produces a manifestation of that disposition. The key causal agent, in this view, is the categorical basis. With the identity between the dispositional and the categorical base secured, it is the *state* of the categorical base that causes the manifestation of the disposition (*MTM*, 88–9; Armstrong 1980, 9; 1984, 22). In this scheme, brittleness would be the state of the microstructure of the brittle object and it is this state that lies behind and causes the object to manifest brittleness—say, the object shattering when dropped in suitable circumstances. In 'The Nature of Mind', Armstrong writes:

> What makes the state a state of brittleness is the fact that it gives rise to the characteristic manifestations of brittleness... If we treat dispositions as actual states... then we can say that dispositions are actual *causes*, or causal factors, which, in suitable circumstances, actually bring about those happenings that are the manifestations of the disposition. A certain molecular constitution of glass that constitutes its brittleness is actually *responsible* for the fact that, when the glass is struck, it breaks. (Armstrong 1980, 8–9)

---

[3] The Lecture was published as a written work and reprinted in 1966. A more fully worked-out account of perception and its relation to dispositionality also appears in Armstrong's 1966 book chapter, 'A Theory of Perception' (Armstrong 1966b), in which dispositions begin to take on implications in terms of causal agents. Revised versions of 'The Nature of Mind' appear in *A Materialist Theory of the Mind* in 1968; and it was re-published in his collection *The Nature of Mind and Other Essays* in 1980.

Note that the categorical state of the microstructure is the disposition and is responsible for any 'causing' going on. Laws of nature had not yet appeared in this discussion on dispositions, and the theory was still silent on *how* the categorical base does the causing or bringing about of the manifestation. And indeed, a further question appears, whether the identification in fact answers Hume's problem of causation. But looking back to this time, in his 1984 self-profile, Armstrong recalls his reasoning as follows,

> It seemed to me, on the contrary [to Ryle's position], that if a dispositional predicate applies to an object, then there must be some actual categorical state of the object in virtue of which the predicate applies, and which, if the disposition should be actualized, would play an essential causal role in the actualization.
> (Armstrong 1984, 22)

In *MTM* Armstrong sets out a more complete defence of his central-state materialism and his identity theory of mind, and in this book he argues at length for his position that dispositions for behaviour are nothing over and above the categorical basis. According to Armstrong, a key difference between the views of Ryle and himself, is that for Ryle there exists at best a contingent connection between dispositions and the categorical bases of objects (or persons) possessing them (*MTM*, 86–7). Since only contingently connected to the microstructure of their bearers, there is no reason why the connection *must* exist in the present, or indeed *would continue* to exist in the future. Also, as Armstrong notes, a further explanation is required to make clear what the connection is, and how and why the connection occurs, an explanation lacking in Ryle's account. Importantly, the contingency of the connection seemed to stand in the way of predictability. Without some firm physical grounding for dispositions, they appeared in Ryle's theory as no more than a linguistic descriptor, and they provided no further illumination concerning the behaviour of their possessors.

In contrast to Ryle's view, Armstrong proposes a reductive account of the connections between dispositions and their bases as necessary, these being a strict (two-way) identity. Armstrong, now writing under the banner of a 'realist' concerning dispositions, writes:

> The Realist view gains some support from ordinary language, where we often seem to identify a disposition and its 'categorical basis'. ('It has been found that brittleness is a certain sort of molecular pattern in the material'.) . . . He [the realist] will say that there is every reason to believe that the categorical state of the band which is responsible for its stretching one inch under force *F* obtains at T_1. Given that it does obtain at T_1, then, as a matter of physical necessity, the band must stretch one inch under force *F*. (*MTM*, 86–7)

There exists a strict identity between a disposition and the categorical base of the object possessing that disposition, the identity constituting a *full reduction* of the disposition to the microstructure of the object bearing the disposition. However, this full reduction includes the additional factor of the state that the microstructure is in. A full reduction is achieved if and when the microstructure is in a certain state that corresponds to the microstructure having a certain disposition. If a glass is fragile, this is because the glass is made of a substance that shatters when in a certain state. The disposition of fragility is identical with the nature of the physical microstructure of the object. However, the physical microstructure must be in a certain state for the fragility of the glass to manifest itself, should the right circumstances arise. The glass must be within a certain temperature range, for example, and made into a particular thickness or shape, etc. That is, the disposition exists along with the state that the categorical microstructure is in. (Note that the 'state of the microstructure' is not explicitly delineated from the 'circumstances of the object in question', at this stage.) Notwithstanding, including the state in which a microstructure is in, the existence of the categorical basis is sufficient to warrant the following conditional claim: If the base were in a certain state (which it currently is not in), then a certain corresponding disposition would exist. Hence, we see that Armstrong's realism about dispositions is both a reductive account (dispositions are identical to the microstructure in a certain state) and also a contingent account (the microstructure must be in a certain state). It is contingent because there is no guarantee that the microstructure will always be in that state, for example, the temperature of glass can change.

Responses to Armstrong's description of dispositions in *MTM* include questions about two possible readings of the identity condition for dispositions: first, the disposition is identical to the microstructure; and second, the disposition relies for its existence on the microstructure being in a certain state (see the discussion that afterward ensued between Armstrong and Squires 1968–1970). If, as per the first condition, the disposition is identical with the categorical basis, then it will always be manifesting so long as the categorical basis exists. But Armstrong's motivation is to allow for the existence of *unmanifesting* dispositions. Squires (1968) claims that for a categorical base to exist without the disposition manifesting, this must mean that the disposition is only a property of the categorical basis, and not strictly identical to the basis itself; dispositions are properties that are acquired when the basis changes its state according to different circumstances. If dispositions are not identified with the base itself, perhaps then they should be identified with the microstructure-in-a-certain-state. However, this would also raise difficulties concerning whether the account can support unmanifesting dispositions. If the disposition only appears with the existence of a certain state of microstructure, then it again must be manifesting when the microstructure is in that state. This critique ties together the existence and manifestation of a disposition, with no option for unmanifested dispositions.

A deeper analysis of these responses is outside the scope of this chapter. However, I note that in much later writing (see, for example, Crane et al. 1996; Armstrong 1997), Armstrong details a division between the existence of and manifestation of dispositions, tying their existence to the microstructure and their manifestation to laws of nature and circumstances. Such a response in the post-*MTM* reviews of his realist account could have possibly answered some of the critical points made by Roger Squires and other reviewers (Armstrong 1969; Goldman 1969; MacIntyre 1968; Nagel 1970; Squires 1968; 1970; Urmson 1969). However, this response was not in close focus in the conversation that followed the publication of *MTM*. Despite this, the realist account of dispositions that is presented in *MTM* played a very important role in opening the way for a much broader discussion on powers, causation, and laws of nature throughout the rest of the twentieth century and into the twenty-first.

4. Conclusion

The very earliest work occupying Armstrong as an undergraduate student began with issues involving topics in the metaphysics of science: causation, necessity, universals and particulars, space and time, continuity, kinds and instances, properties and relations, distinctness, particulars, and powers. Key themes included the denial that necessary connection is required for uniformity and that spatio-temporal relations, together with concepts of sorts and kinds, are sufficient to account for the patterns of regularity and causation.

From around 1952 to 1954, Armstrong's focus begins to turn toward the topic of perception, and from there his work portrays a consistent effort to defend and develop the direct realist theory of perception. The birth of his focus on dispositions arises as his earlier work in causation overlaps this work in perception. The central-state theory of the mind begins to overlap with questions about qualities and properties, individuation, instantiation, and causation, and from there to an examination of the link between dispositions and mental states. It is around the early 1960s, after his publication of *Perception and the Physical World* (Armstrong 1961), that Armstrong begins his journey proper into his investigation into dispositions and produces the work that will draw together his first comprehensive account of them, this account featuring as a significant component in the arguments presented in *MTM*.

Key influences in the development of Armstrong's realist account of dispositions include Gilbert Ryle, who was a key player in providing Armstrong with incentive for the development of a counter-position on dispositions. Charlie Martin was also an important influence in his discussions with Armstrong concerning the nature of dispositions and their relation to and role they play in mental events. These influences, together with the constraints of Armstrong's strong naturalist stance, produce a shift in Armstrong's position on dispositions from a traditional Rylean view to a realist account in which dispositions are part of the physical world and are identified with the physical microstructure of their bearers. Part of the rationale behind this change to a realist view of dispositions is to provide a warrant or justification for dispositional ascriptions in terms of their causal roles, and in particular, their roles in experiential situations, such as the experience of pain. Talk of causal roles now becomes talk of how the categorical base produces the manifestation of dispositions. Dispositions are identified with the microstructure of the disposition bearer, with the caveat that dispositions are identified with the state of the microstructure in a certain state.

Critiques of Armstrong's realist account of dispositions include questions concerning whether dispositions are to be identified with the microstructure itself or with the microstructure as it stands in a certain state. Either way, it seems that the underlying motivation for putting forward the realist account of dispositions is frustrated by the fact that the account is unable to convincingly account for unmanifesting dispositions. Nonetheless, the realist account of dispositions achieved a great deal by opening the philosophical debate and discussions around the role of dispositions in causation, laws of nature, and as powers.

References

Note: All unpublished manuscripts and correspondence cited in this chapter are housed in the National Library of Australia, Special Collections, Canberra, under the title, MS9363 Papers of David Armstrong.

Armstrong, D. M. (1949a) 'Nature of Belief', unpublished manuscript, MS9363, Series 1, Box 1, Folder 4.

Armstrong, D. M. (1949b) Science as Superstition', unpublished manuscript, MS9363, Series 1, Box 1, Folder 4.

Armstrong, D. M. (1950) 'Honours Thesis: A Realist Account of Kant's "Transcendental Analytic"', unpublished manuscript, MS9363, Series 1, Box 1, Folder 6.

Armstrong, D. M. (1951) 'The Meaning of Free Thought', unpublished manuscript, MS9363, Series 1, Box 1, Folder 7.

Armstrong, D. M. (1952a) 'Absolute and Relative Motion', unpublished manuscript, MS9363, Series 1, Box 1, Folder 8.

Armstrong, D. M. (1952b) 'Criticism of J. Mackie's Paper on Extension', unpublished manuscript, MS9363, Series 1, Box 1, Folder 8.

Armstrong, D. M. (1952c) 'Is Position in Space and Time Absolute or Relative?', unpublished manuscript, MS9363, Series 1, Box 1, Folder 8.

Armstrong, D. M. (1952d) 'On Points', unpublished manuscript, MS9363, Series 1, Box 1, Folder 8.

Armstrong, D. M. (1952e) 'Space-Time-Supplementary Material', unpublished manuscript, MS9363, Series 1, Box 1, Folder 8.

Armstrong, D. M. (1953a) 'Identity of Indiscernibles', unpublished manuscript, MS9363, Series 1, Box 2, Folder 11.

Armstrong, D. M. (1953b) 'Is Hume's Analysis of Necessary Connection in Terms of a Determination of the Mind Circular?', unpublished manuscript, MS9363, Series 1, Box 2, Folder 12.

Armstrong, D. M. (1953c) 'Mill on Liberty', unpublished manuscript, MS9363, Series 1, Box 1, Folder 9.

Armstrong, D. M. (1953d) 'The Nature of Pain', unpublished manuscript, MS9363, Series 1, Box 1, Folder 8.

Armstrong, D. M. (1954a) 'Bachelor of Philosophy Thesis: The Nature of Complexity', unpublished manuscript, MS9363, Series 1, Box 2, Folder 13.

Armstrong, D. M. (1954b) 'Infinite Divisibility of Time', unpublished manuscript, MS9363, Series 1, Box 2, Folder 15.

Armstrong, D. M. (1954c) 'Berkeley's Puzzle about the Water that Seems Both Hot and Cold', *Analysis*, 1(2): 44–6.

Armstrong, D. M. (1955) 'Illusions of Sense', *Australasian Journal of Philosophy*, 33(2): 88–106.

Armstrong, D. M. (1956a) 'Discussion: Berkeley's *New Theory of Vision*', *Journal of the History of Ideas*, 17(1): 127–9.

Armstrong, D. M. (1956b) 'Lecture Series: Kant's Critique of Pure Reason', unpublished manuscript, MS9363, Series 2, Box 10, Folder 4.

Armstrong, D. M. (1956–1958) 'Lecture Series: Idealism and Realism (George Berkeley's Idealism and Realist Critiques of F. H. Bradley's Absolute Idealism)", unpublished manuscript, MS9363, Series 2, Box 10, Folder 5.

Armstrong, D. M. (1957–1958) 'Are the Primary Qualities Really Primary?', unpublished manuscript, MS9363, Series 1, Box 2, Folder 17.

Armstrong, D. M. (1958a) 'The Location of Pain', unpublished manuscript, MS9363, Series 1, Box 2, Folder 19.

Armstrong, D. M. (1958b) 'The Nature of Perception', unpublished manuscript, MS9363, Series 1, Box 3, Folder 20.

Armstrong, D. M. (1960a) *Berkeley's Theory of Vision: A Critical Examination of Bishop Berkeley's Essay Towards a New Theory of Vision*, Melbourne: Melbourne University Press.

Armstrong, D. M. (1960b) 'Perception and the Physical World', unpublished manuscript, MS9363, Series 1, Box 3, Folder 22.

Armstrong, D. M. (1961a) 'Causality', unpublished manuscript, MS9363, Series 1, Box 3, Folder 26.

Armstrong, D. M. (1961b) *Perception and the Physical World*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1961c) 'Mind and Consciousness', unpublished manuscript, MS9363, Series 1, Box 3, Folder 23.

Armstrong, D. M. (1961d) 'Mental Images', unpublished manuscript, MS9363, Series 1, Box 3, Folder 25.

Armstrong, D. M. (1962) *Bodily Sensations*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1963a) 'Absolute and Relative Motion', *Mind*, 72(286): 209–23.

Armstrong, D. M. (1963b) Armstrong to J. T. Stevenson, 1 December 1963, MS9363, Series 1, Box 4, Folder 31.

Armstrong, D. M. (1965) 'Inaugural Lecture: The Nature of Mind', unpublished manuscript, MS9363, Series 1, Box 4, Folder 34. Delivered 10 May 1965.

Armstrong, D. M. (1966a) 'The Nature of Mind', *Arts* (Proceedings of The Sydney University Arts Association), 3(1): 37–58.

Armstrong, D. M. (1966b) 'A Theory of Perception', in *Scientific Psychology*, edited by B. Wolman, New York: Basic Books, 489–505.

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1969) 'Dispositions are Causes', *Analysis,* 30(1): 23–6.

Armstrong, D. M. (1980) 'The Nature of Mind', in *The Nature of Mind and Other Essays*, St Lucia: University of Queensland Press, 1–15.

Armstrong, D. M. (1983) *What Is a Law of Nature?*, Cambridge: Cambridge University Press.

Armstrong, D. M. (1984) 'Self-Profile', in *D. M. Armstrong, Profiles*, vol. 4, edited by Radu J. Bogdan, Dordrecht: D. Reidel, 3–51.

Armstrong, D. M. (1997) *A World of States of Affairs*, Cambridge: Cambridge University Press.

Ayer, A. J. (1956) *The Problem of Knowledge*, London: Macmillan.

Berkeley, George (1949) *The Works of George Berkeley Bishop of Cloyne*, 9 vols, edited by A. A. Luce and T. E. Jessop, London: Thomas Nelson and Sons. [D. M. Armstrong annotated various volumes in this edition.]

Carnap, Rudolph (1936) 'Testability and Meaning', *Philosophy of Science*, 3(4): 419–71.

Choi, Sungho and Michael Fara (2006) 'Dispositions', in *Stanford Encyclopedia of Philosophy*, Stanford, CA: Stanford University. https://plato.stanford.edu/

Crane, Tim, D. M. Armstrong, U. T. Place, and C. B. Martin, eds (1996) *Dispositions: A Debate.* New York: Routledge.

Goldman, Alvin I. (1969) 'Review of D. M. Armstrong, *A Materialist Theory of the Mind*', *Journal of Philosophy*, 66(22): 812–18.

Heil, John (2010) 'Martin, C. B.', in *A Companion to Philosophy in Australia and New Zealand*, edited by Graham Oppy and N. N. Trakakis, Melbourne: Monash University Publishing.

Hume, David (2000) *An Enquiry Concerning Human Understanding*, edited by Tom L. Beauchamp, Oxford: Clarendon Press.

Hume, David (2007) *A Treatise of Human Nature*, 2 vols, edited by David Fate Norton and Mary J. Norton, Oxford: Clarendon Press.

Kant, Immanuel (1950) *Immanuel Kant's Critique of Pure Reason*, trans. Norman Kemp Smith, London: Macmillan. [D. M. Armstrong annotated his copy of this edition.]

Kant, Immanuel (1997) *Critique of Practical Reason*, translated by Mary J. Gregor, Cambridge: Cambridge University Press.

Kant, Immanuel (2000) *Critique of the Power of Judgment*, translated by Paul Guyer, Cambridge: Cambridge University Press.

Locke, John (1975) *An Essay concerning Human Understanding*, edited by Peter H. Nidditch, Oxford: Clarendon Press.

MacIntyre, Alasdair (1968) 'Materialism Revisited: A Materialist Theory of the Mind (D. M. Armstrong)', *New Society*, 11(279).

Martin, C. B. (1957) 'Dispositions and Conditionals', unpublished manuscript, (MS Acc12.001, Box 27, Folder Martin.

Martin, C. B. (1958) 'Communication', MS9363, Series 1, Box 2, Folder 19.

Martin, C. B. (1994) 'Dispositions and Conditionals', *Philosophical Quarterly*, 44(174): 1–8.

Nagel, Thomas (1970) 'Armstrong on the Mind', *The Philosophical Review*, 79(3): 394–403.

Ryle, Gilbert (1949) *The Concept of Mind*, London: Hutchinson. [D. M. Armstrong annotated his copy of this edition.]

Squires, Roger (1968) 'Are Dispositions Causes?', *Analysis*, 29(1): 45–7.

Squires, Roger (1970) 'Are Dispositions Lost Causes?', *Analysis*, 31(1): 15–18.

Urmson, J. O. (1969) 'Review of D. M. Armstrong, *A Materialist Theory of the Mind*', *Annals of Science*, 25(1): 83–4.

5

Materialism, Then and Now

Peter Godfrey-Smith

1. Introduction

David Armstrong was the Challis Professor of Philosophy when I arrived as a first-year undergraduate at the University of Sydney in 1983.[1] He taught epistemology in first term. (Bill Lycan, visiting from the USA, handled metaphysics). I was told that Armstrong taught first years regularly because he thought that all students should be taught at some stage by 'The Professor'. The lectures were good—patient, considered, and fair. They also encouraged us to see philosophical debates in rather competitive, combative terms ('spoils to the victor', 'a problem for everyone is a problem for no one'—I am not sure if I heard those phrases in that first class, but certainly early on). In my fourth year he taught a whole-year course that was essentially the writing of *A Combinatorial Theory of Possibility* (Armstrong 1989). I was not entirely on board with that project and its methods, but enjoyed the course a great deal and relished, as always in the T&M department in those days, the sense that I was seeing philosophy that mattered being developed and shaped in real time.

The T&M or 'Traditional and Modern' Philosophy Department in which Armstrong taught was one of two at Sydney, born of a split in the 1970s. During my time at Sydney in the 1980s, a lot of work in the philosophy of mind was being done within a more American style, one imported in part by Michael Devitt, who had returned from Harvard some time before, a style fostered also by American visitors like Bill Lycan and Stephen Stich. We talked about mental representation, intentionality, and computation, reading Jerry Fodor and Dan Dennett (and in my case, Fred Dretske and Ruth Millikan). Armstrong seemed to some of us a bit stodgy in this context, and he did push back against that. His annoyance was sometimes evident, especially at the idea that there was a history that ran from 'the identity theory' to 'functionalism', with him as the last part of the first stage. He denounced 'trendiness' in naturalistic philosophy. He was,

---

[1] The text here is fairly close to the talk given at the Armstrong conference in August 2018, but I have filled out the last part of each of the main sections, and incorporated a couple of points made in discussion. Other points have been amplified in footnotes. Thanks to Peter Anstey and David Braddon-Mitchell for organizing the conference, and for helpful comments on an earlier draft.

Peter Godfrey-Smith, *Materialism, Then and Now* In: *Armstrong's Materialist Theory of Mind.* Edited by: Peter R. Anstey and David Braddon-Mitchell, Oxford University Press. © Peter Godfrey-Smith 2022.
DOI: 10.1093/oso/9780192843722.003.0006

again, a combative sort of person. This affinity with conflict served him in good stead during the creation of the T&M Department, which was, as David Stove always said, a wonderful place to be. It enabled Armstrong to get through a period in the 1970s when he was actively shunned and excluded from social events—with groups leaving lunch tables at the staff club when he sat down.[2] Some of those people later benefited greatly from the birth of the T&M Department, and I hope they apologized.

Armstrong was in other ways modest. I heard him say more than once that his hope was to be a footnote to the history of philosophy. I think it is becoming clear that it is only in the briefer histories that he is just a footnote. He also endorsed the 'under-laborer' description of philosophy, which was surprising to me. Perhaps he thought that if this was good enough for Locke, it was good enough for him. But this is obviously inaccurate, as is the official emphasis on conceptual analysis in *A Materialist Theory of the Mind*. The book is full of material that is more substantive, more abstract-empirical, as he later admitted in the Preface to the paperback edition.

My first topic in this chapter is how Armstrong set up his materialism, especially in *A Materialist Theory of the Mind* (hereafter *MTM*). Then I'll say something about how we should think about materialism now.

2. Materialism Then

Armstrong's view is expressed as a 'causal theory', a causal analysis of mental states and processes. It is a view that grants quite a lot to behaviourism, while holding that mental states are inner states that act as causes. We get a handle on mental states via their causal role, especially as causes of behaviour. Science can then tell us that particular neural states and processes are the things that do that causing, the things that play these roles in the actual world. As a result, we can identify mental states and processes with those neural states and processes.

> The concept of a mental state is primarily the concept of *a state of the person apt for bringing about a certain sort of behaviour*. Sacrificing all accuracy for brevity we can say that, although the mind is not behaviour, it is the *cause* of behaviour. In the case of some mental states only they are also *states of the person apt for being brought about by a certain sort of stimulus*. But this latter formula is a secondary one. (*MTM*, 82)

---

[2] I learned this recently from Kim Sterelny.

Two things he does not worry much about are the modal status (necessity/contingency) of his claims, and multiple realizability. In the paperback Preface he says he accepts multiple realizability and always did, once the question was raised.

In retrospect, I think we can see Armstrong's project as follows. He wants to cover the whole of the mind (near enough), but he does this by working his way in from two particular parts of the mind that get a great deal of attention. These are *perception* (and its relatives), and states related to the *will* (such as intentions). Both these aspects of the mind are *peripheral,* in a sense. When I say that, I don't mean they are secondary in importance, but peripheral in a more literal sense. They are close to the mind's interaction with the non-mental. Armstrong does not present them this way, or give this as the reason for his focus on them. He just says they are important. He says more than once that perception and the will are 'great' and 'central' topics for the philosophy of mind.[3] He sees himself as going after what is basic. But we can also see their peripheral character, on each side. These peripheral states can be contrasted with what I will call the *core*. The core is belief-desire interaction—the way that actions stem from what you want in the light of what you believe—together with the role of memory and other faculties that bear on belief-desire interaction. Perception feeds beliefs, and the belief-desire nexus gives rise to intentions and actions.

The shape of belief-desire interaction, and how to model it rigorously, were seen clearly by Ramsey in 'Truth and Probability' (written in 1926)—perhaps by others even earlier, but certainly by Ramsey. I will say more about Ramsey's achievements in relation to Armstrong below.

In the light of Armstrong's emphasis on perception and intention, we can see him as situated between and responding to British empiricism, on one side, and, on the other side, Rylean behaviourism and Wittgensteinianism. The empiricist influence is clear: Locke and Hume are often discussed, and sometimes also Berkeley. But Armstrong was also impressed with Ryle, and the achievements of the people around him and Wittgenstein. *MTM* includes lavish praise for Wittgenstein and behaviourism—talk of 'astonishing progress' and 'epoch-making philosophical achievements'. Whatever the final view of the mind is, Armstrong says, it must make peace with behaviourism.[4]

He took seriously both of these very different views: *Empiricism*, with its individualist and first-person character, with its proneness to scepticism and phenomenalism, and its running into trouble with more cognitive aspects of the mind; and *behaviourism*, with its social, third-person viewpoint, and its deflationary outlook on the tradition of sensationalist empiricism. Armstrong was influenced by both, and each has their favourite mental states: perception on the empiricist side, and the states most closely related to action, such as intention, on the other side.

---

[3] See (*MTM*, 187 and 245).

[4] See (*MTM*, 67 and 118).

A consequence of this approach, though, is a failure of a causal theory of the mind of Armstrong's kind. The core, the belief-desire nexus, resists an analysis of the sort he gives, and resists it in part *because* it is non-peripheral. I'll discuss this point using the case of belief, first through a direct application of Armstrong's summary above, and then with a look at what Armstrong explicitly says about belief in a late chapter.

Suppose we apply Armstrong's formula directly. If you have a particular belief, you are disposed to do certain things; there are behaviours the belief is apt to cause. This would be a modification of the Rylean idea that to believe something is to have (complex) dispositions to behave in certain ways. The problem is that the effect on action of any belief depends on other beliefs and on desires. Two people who both believe that Baltimore is in Maryland could behave indefinitely differently, depending on what they want and what else they believe. Even if they are asked a yes-or-no question in a quiz show with $1 million at stake, they will answer differently if one thinks they are being tricked, or that the money will corrupt them, or whatever. The thing that causes an action is a complex of beliefs and desires.

Armstrong does not want to be simplistic about the causal role of core states. He looks for ways to hedge and qualify what is claimed. That would be OK if there was a central tendency in causing behaviour that each belief had, one that might get interfered with in various ways. But there is no central tendency in a belief's role. Core states are just not amenable to a treatment of this kind. The idea of '*an* inner state bringing about outward behaviour' (*MTM*, 80, emphasis added) just does not apply to beliefs.

There is a standard objection to Ryle and logical behaviourism called the 'Chisholm-Geach' objection, which holds that attempts to describe the dispositions associated with a particular mental state will collapse into circularity, because the behaviours associated with the belief that *p* will depend on other mental states. As Lycan noted in discussion, my point here is that the Chisholm-Geach objection is as much an objection to Armstrong as to behaviourism.[5]

---

[5] See Chisholm (1951) and Geach (1957). Here is Geach:

> [I]s there in fact any behaviour characteristic of a given belief? Can action be described as 'acting as if you held such-and-such a belief' unless we take for granted, or are somehow specially informed about, the needs and wants of the agent? In Ryle's example this information is smuggled in by his speaking of a *gardener's* rain-expecting behaviour (and tacitly assuming that the gardener is not e.g. a discontented or corrupt servant who wants the garden to be ruined). When Dr Johnson did penance in Uttoxeter market-place, he may have begun by standing around bareheaded until the threatened shower should fall; this would not be recognizable as rain-expecting behaviour without a knowledge of Johnson's wish to do penance. (Geach 1957, 8)

Armstrong's view is close enough to behaviourism to inherit one of its main problems.

In understanding core states, there is no way of getting by with less structure than a belief-desire profile, of the sort Ramsey worked with in 'Truth and Probability'. Here is Ramsey:

> I propose to take as a basis a general psychological theory, which is now universally discarded, but nevertheless comes, I think, fairly close to the truth in the sort of cases with which we are most concerned. I mean the theory that we act in the way we think most likely to realize the objects of our desires, so that a person's actions are completely determined by his desires and opinions. This theory cannot be made adequate to all the facts, but it seems to me a useful approximation to the truth particularly in the case of our self-conscious or professional life, and it is presupposed in a great deal of our thought. It is a simple theory and one which many psychologists would obviously like to preserve by introducing unconscious desires and unconscious opinions in order to bring it more into harmony with the facts. (Ramsey 1931, 173)

That move is not yet pushing us far from an Armstrong-type view. Something that does push away from it, however, is Ramsey's use of an expected utility model, again as an approximation: you will do what is associated with the best expected payoff, all things considered. That requires that degrees of belief be recognized; this is part of what makes it possible to use very small differences in behaviour (such as whether you would walk 100 yards or 120 yards out of your way to ask directions) to get a handle on mental states. Behaviour stems from degrees of belief and from evaluations of events that are the possible outcomes of actions.[6] (Again, Ramsey keeps saying this is an approximation, but the ways in which it is an approximation do not take us back toward an Armstrong-like view.) If this model is accepted, what remains of belief as 'an inner state bringing about outward behaviour'? A (degree of) belief might be a function (in the mathematical sense) from contexts to behavioural dispositions, where each 'context' is much of the rest of a psychological profile. That is the closest a belief can get to being disposed to issue in some particular behaviour.

What about desires? Do they have the same 'core' properties? Is it true that a desire can issue in almost any behaviour, depending on what is believed and what else is wanted? Armstrong has a long discussion of forms of desire that are removed from causing acts directly, but action-causing remains central to his view: 'Desires are essentially action-producing although they may be inhibited or prevented from producing action by other desires' (*MTM*, 152). Are there some desires that are 'apt for causing' actions in a straightforward way? Or are the effects of all desires mediated by beliefs, with the same holistic features I associate with the core and the Ramsey model?

---

[6] According to Ramsey (1931, 173):

> I suggest that we introduce as a law of psychology that his behaviour is governed by what is called the mathematical expectation; that is to say that, if p is a proposition about which he is doubtful, any goods or bads for whose realization p is in his view a necessary and sufficient condition enter into his calculations multiplied by the same fraction, which is called the 'degree of his belief in p'. We thus define degree of belief in a way which presupposes the use of the mathematical expectation.

I think desires are a bit special. When desires are directed at outcomes distinct from acts, then their role is always indirect and belief-mediated. I desire to get to work on time. That could result in almost any act, depending on my beliefs. But a few desires seem to be aimed at acts alone. I desire to shut my eyes (not: that my eyes end up closed). This desire does not apparently have the holistic features. 'Desire' then covers both core and non-core states.[7]

I said above that the core was described by Ramsey, decades before Ryle and Armstrong. Ramsey was remarkably philosophically independent. You might say that he was not around long enough for the traditions to seep into him, but he did go quite far down his own distinctive path. To a significant extent, Ramsey made the core unmysterious, by showing how subtle differences between belief-desire profiles can be made manifest. I have to say all this in a qualified way; Ramsey did not give a behaviourist reduction of the core. The bridging components in his analysis are dispositions to choose between big hypothetical scenarios, and between gambles over those scenarios. (He says things like this: 'If then we had the power of the Almighty, and could persuade our subject of our power, we could, by offering him options, discover how he placed in order of merit all possible courses of the world' (Ramsey 1931, 176). As Ed Elliott (2015) discusses, this does not really forge a constitutive link to behaviour. The same holistic factors discussed above arise for behavioural expressions of these preferences, too. The subject's mental 'placing in order' and their behavioural 'placing in order' need not simply align. We might understand the Ramsey project as using an idealization, making use of a peripheral mental state that has special properties (reminiscent of Armstrong), or the project might be carried forward differently. But a lot of progress is made in handling the behavioural manifestation of core states.[8]

Armstrong—like nearly everyone, with the partial exception of Lewis—was not looking to Ramsey *in this way*. Not even Ryle was.[9] It is odd how little connection there was, for a long time, between Ramsey's discussion and the rest of philosophy. Lewis in 'Psychophysical and Theoretical Identifications' (Lewis 1972) gave an account that can apply to the core. That account used the idea of a 'Ramsey sentence', which is derived from a different Ramsey paper.[10] In his 1972 paper, Lewis did not discuss *how* belief-desire profiles might be made manifest and did not note what Ramsey had done in 'Truth and Probability'. He just discussed the general way that Ramsey sentences might be used to make sense of a complex network of causal roles and the occupants of those roles.

---

[7] This might be called an 'intention' rather than a desire. I think all the language around here is pretty flexible.

[8] Elliott (2015) looks at where and why various other attempts fall down on the most ambitious task here, showing how pure behavioural dispositions might reveal preferences and degrees of belief. Elliott's treatment has a glass-half-empty character, which should not be seen as the whole story, though I concede that I don't know how to get more into the glass.

[9] Ramsey is not in Geach's (1957) book either. There's a question about the history of the field that would be worth a close look here. There seem to be some sort of subtle but powerful framing effect at

I now return to Armstrong. Armstrong discusses belief at various points in the central chapters of *MTM*, but as he is focused so much on perceptions and intentions in those chapters, belief is not handled in detail. Then, in a short late chapter, he addresses belief specifically. Here he says that *perceptual beliefs* are basic, and those are understood in terms of their enabling discriminative behaviours, and also says that perceptual beliefs (taken together) are like a big 'map' of the environment. *Non*-perceptual beliefs (beliefs held today about Caesar, and so on) are *extensions* of the map. Those beliefs are 'a "map" of reality that is an extension of the "map" formed by our perceptual beliefs' (*MTM*, 343). What sort of extension? As he says both in 1968 and the later paperback Preface, his discussion here is very thin. Suppose one took it seriously and as literally as seems possible. Then, insofar as our beliefs are map-like, they would affect action in something like the way seen in Ramsey's treatment: actions are determined by the big map (with confidence levels) together with evaluations of the likely effects of actions. We would end up far away from the idea that beliefs are apt for causing particular behaviours. The 'map' idea is taking us down a different path.[11]

Armstrong at other places explicitly tries to qualify his 'apt for causing' formula in the case of things like beliefs. A perceptual belief does not really issue in particular behaviours, he says, but *makes possible* some particular behaviours, which are discriminative behaviours. You might not want to perform those behaviours, but you can if you want:

> [I]n the case of perception, there is no question of the inner event actually tending to bring about behaviour. What we must say, rather, is that perception supplies a *necessary precondition* for appropriate behaviour. If the baby can perceive the difference between a green and a blue block, then it is in a position to discriminate between them in its behaviour *if it should want to*... Perception can then be compared to acquiring a key to the door. If I acquire such a key,

---

work, where Ramsey's discussion is often not perceived as relevant. I was not thinking about it in the present connection until I looked at the connections between Peirce and Ramsey, and then encountered Elliott's work.

[10] Apparently the importance of this idea was recognized first by Hempel. Here I draw on Psillos (2004), who says that Hempel coined 'Ramsey-sentence' in his 1958 'The Theoretician's Dilemma', drawing on Ramsey's posthumous 1929 'Theories'. Psillos says that Braithwaite had used Ramsey's paper a bit, and Carnap had reinvented the idea before he read Hempel's paper.

[11] This path is explored in some detail by Braddon-Mitchell and Jackson (2006).

> I have acquired an instrument which may be a necessary precondition for the opening of the door, even if I subsequently ignore the door completely.
> (*MTM*, 249)

However, to say this is to completely abandon the apt-for-causing analysis, the approach guided by ideas like this:

> if we say that the baby's perception is the coming to be of a state of the baby apt for the bringing about of certain sorts of discriminative behaviour, we get the best of both realms. We preserve the 'inner' character of perception, yet at the same time we create a logical tie between the inner event and the outward behaviour.
> (*MTM*, 248)

Armstrong may have thought that if we drop that last idea, we are doomed; we have to hang onto some version of that last idea in order to give any causal analysis of the mind at all. But Ramsey had made progress on this problem decades before. He showed that if you do the *double* complication—from simple belief to degrees of belief, and from individual states to whole profiles—you can go a long way toward making sense of the behavioural manifestation of core psychological states. Those complications make things better, not worse.[12]

Before moving on, I'll say something about Armstrong's treatment of belief in the most relevant work after *MTM*, which is *Belief, Truth and Knowledge* (Armstrong 1973). This book expands on the 'map' analogy for beliefs, and this involves yet another distinct link to Ramsey. Ramsey, in 'General Propositions and Causality' (Ramsey 1990), talked of belief as comprising a 'map . . . by which we steer'. In the 1973 book, Armstrong takes this on board, and accepts, as I read him, that this moves us away from the kind of causal analysis given in *MTM*.[13]

On this new road, different questions arise. Map-talk might be understood here as referring to the belief side of the belief-desire nexus as handled in an expected utility model. That leaves open the possibility of a holistic determination of the belief-desire profile of an agent at a time. Alternatively, the 'map' idea could point toward inner states with a particular kind of combination of representational and causal properties—picture-like representations that we use to help us steer.

---

[12] Another way to put it: without degrees of belief, the 'Chisholm-Geach' phenomenon is just a problem, something that messes up various reasonable-looking projects. With degrees of belief, which Armstrong and others set aside as a secondary matter, the core becomes something that can be represented and studied. What makes things look harder ultimately makes them easier.

[13] According to Armstrong (1973, 71):

> It is an essential mark of beliefs, then, as opposed to mere thoughts, that, if suitable dominant desires are also present, the believer is moved to action: action having as its objective the satisfaction of the desire. This does *not* mean that an account can be given of beliefs as being nothing but causal factors in the believer which initiate and sustain certain courses of action in certain circumstances.

Perhaps there is a place for both. We are thus drawn into the kinds of debates about intentional psychology and mental representation seen in the 1980s. Are there 'explicit' mental representations, and how do they work? The idealizations made in an expected utility model now start to matter. Which features of ordinary belief-desire interpretations are aimed at picking out inner causes, and which ones have a 'looser fit' to what is going on inside?[14]

3. Materialism Now

The second half of the chapter will look at different themes: at what materialism *is*, and why we should believe it.

What is it to be a materialist? Armstrong hardly discusses this question in *MTM*, taking it for granted that if you identify mental states with 'physico-chemical states of the central nervous system' (*MTM*, 356), then you are a materialist, at least about the mind. But recent discussion has shown that none of this is straightforward. The *material* or the *physical* is not a very clear category. (I will often use the term 'physicalism' from here, but I don't intend this as a significant shift.[15]) We shouldn't commit to materialism or physicalism by saying that textbook physics in its current form gives us an inventory of everything that exists, or the low-level basis of everything that exists. Present-day physics is probably not entirely accurate. It can be tempting to say that to be a physicalist is to think that a future and completed physics will describe everything that exists (or its low-level basis), but that cannot be the right approach, either. If it turns out that God, or a World Spirit, exists, the Physics Department can be expected to take an interest. A future physics might not be physicalist, in the sense on the table here. Perhaps a physicalist is committed to the idea that the physics of the future will be a clear descendant of current views, which are on the right track? This is what Lewis (1994) said. I think this is wrong, too. We can be confident, perhaps, that future physics will agree with present physics about a large range of macroscopic phenomena, but are we so sure it will agree on what sorts of things are fundamental? Lewis said: 'We may think, for instance, that mass and charge are among the fundamental properties; and that whatever fundamental properties remain as yet undiscovered are likewise instantiated by very small things that come in very large classes of exact duplicates' (Lewis 1994, 51–2). This is too strong. Suppose the physics of the future uses a field ontology rather than a particle ontology and treats particle-like phenomena as derivative. That is not very far-fetched—various physicists have believed it. It must be possible for a future field-based view to count as physicalist.

---

[14] This last idea is followed up in detail by Dennett in his neo-Rylean works; see, Dennett (1971; 1987).

[15] Lewis (1994) says that it would be pedantry, or worse, to switch from 'materialism' to 'physicalism'. Both terms do have problems. A modern version of 'materialism' is as much about energy as matter, as Lewis notes, and 'physicalism' suggests a relationship to a field of study rather than a kind of entity, something I think would be a mistake.

How do we steer through this? Is there a view of materialism or physicalism that is broad enough to allow surprises that involve big holistic fields, but narrow enough to rule out a future physics of spirit? We should accept that the category of 'the physical' has become a bit unclear, but I think there is a way to steer through. The heart of my own view is: (1) monism, and (2) the construction of mind from things that are not mental (and have no mental properties). This leaves some flexibility about what counts as a physical fact, but does not fall into panpsychism. This view is related to what is now sometimes called *via negativa* physicalism. I'll discuss toward the end of the chapter whether this is strong enough to mark out materialism from all other options; I think some of the boundaries are a bit vague.[16]

Why should one believe a view of this general form? I am going to take it as a background assumption that standard anti-materialist arguments have adequate replies. Here I have in mind the 'knowledge' argument, zombie arguments, and so on. Some of the standard replies are adequate in each case.[17] My main topic in this section is the positive arguments that can be given for materialist views.

A good way to make an argument of this kind is to make use of continuities between us and other organisms, with respect to the relations between physical and mental properties, and use these continuities as a constraint on overall views. I don't make the argument using parsimony. Parsimony considerations are important here for many philosophers, including Smart (1963) and Armstrong, but I regard parsimony arguments, in their (slender) reasonable forms, as inapplicable in this context. Instead, we should ask: what are the *rival overall pictures*, and which ones seem likely to be filled out in a way that works, especially when we have our eye on all sorts of living things? Dualism is most plausible when we can set things up in a way that contrasts humans with everything else. Descartes was able to use a framing of this kind. It is much harder to be a dualist when you accept psychological continuities extending through many kinds of organisms and an evolutionary story that links all these features. It is hard to be a dualist when you want to make sense of continuities between human mental phenomena, non-human animal mentality, and the proto-cognitive capacities seen in all cellular life.

---

[16] See Montero and Papineau (2005) for a discussion of *via negativa* options.

[17] For knowledge arguments, see Jackson (1982), and for zombie arguments, see Chalmers (1996). Godfrey-Smith (2019) discusses what I see as the right replies in each case.

I will develop that argument further in a moment. First, I'll say something about the other approach, which uses parsimony. Parsimony arguments are prominent in Armstrong's book. They are invoked often in the final chapter, where Armstrong wraps up his case for materialism, especially when he has to deal with non-classical dualist views. For many, the problem for traditional dualist views is the problem of interaction, but weaker dualisms are harder to rule out, including epiphenomenalism and emergentism of some kinds. Parsimony deals with those.

The argument from parsimony is offered, by Armstrong and others, in a setting in which it is unclear whether there is *any* reason that simplicity is a guide to truth, on this issue or others. No positive reason is offered to believe that simplicity is a guide to truth. This is surely a major problem. What lies behind many uses of parsimony of this kind is perhaps the implicit message that if this is not reasonable, then all hell breaks loose—not just here, but everywhere. Abandoning parsimony leads to a complete falling-apart of rational theory choice, as there will always be many options compatible with our evidence. Parsimony about the mind-body problem, then, might be an application of a principle that might be hard to justify, but one that we can't possibly do without.

A general rethinking of simplicity preferences is needed, both in science and philosophy. Simplicity doesn't mean nothing, but it's not a guide to truth, especially in cases like this. A brief sketch of the view I'd defend is as follows. Simplicity does not give us reason to believe a theory is true, but simplicity preferences do have a role in science. First, a preference for simplicity is part of a good policy or strategy aimed at scientific change. A good rule is to start simple and expect to get pushed elsewhere. This is a broadly Popperian attitude. Suppose instead we began with a more complex theory. It is no less likely to be true than the simple one, but the process of being pushed from old to new views by incoming data is less straightforward. Simple theories are good places from which to initiate the dynamic process characteristic of theory development in science. Occasionally a simple theory might actually be true; that is merely a bonus.

Second, there can be domain-specific empirical reasons to prefer simplicity in some cases (Sober 1990). If you think that beneficial mutations are rare events in evolution, then a story that requires fewer of them has an advantage. Third, in a Bayesian context, the Popperian point above has a positive flipside, one sometimes now known as 'Bayesian Occam's razor' or the 'automatic Occam's razor'. When a theory constrains the data a lot (takes risks, is easily falsified), if incoming data *does* fit the theory, the theory gets a lot of credit, more than it would if it had only weakly constrained the data. In some cases, simple theories do constrain the data more than related theories that are more complex. In other cases, they don't—a simple theory might also be vague about some of what is going on. Simplicity in this Bayesian context does not operate as an extra principle that gives one theory an edge over others after the data has done its job. Instead the razor, when real, is a reflection of the fact that theories that constrain the data a lot get credit when their predictions are borne out, and in some cases, simple theories constrain the data a lot.[18]

None of these three scientific roles for simplicity applies to the mind-body problem.

If you want more than this from parsimony, then tell me how we can have more. Don't say: if there is no more than this, then we're doomed. We're not doomed in ordinary scientific contexts, where we are continually getting new data, and not doomed either in the philosophical cases. In that context, our project should be to get a sense of the range of genuine options, and choose the best one according to how all these options handle various kinds of evidence. The problem with many alternatives to materialism is not that they are unparsimonious, but they don't have resources with which to tell a coherent story. When they are fleshed out, they do badly with what we know.[19]

Here is a more detailed outline of an argument of that kind, using the continuities between human and non-human organisms. The starting point is our own case—our own brains, the cognitive side of the mind, and subjective experience. There does seem to be an initial difficulty in explaining experience (consciousness, in a minimal sense) using those resources, so dualism becomes a possibility. Then we look at other animals, finding not just brains and the control of behaviour, but many apparent marks of experience. We have an outline of an evolutionary story about how these capacities arise, including ones that seem to have some relation to felt experience. We can trace the evolution of sensing and the perception of objects, of the internal valuation of events, of the ongoing tracking of the divide between self and other, which is needed by a mobile animal to make sense of what it perceives (Merker 2005). All these capacities have simple beginnings—some in unicellular life, some in invertebrate animals—and there is no reason to believe in an evolutionary leap in the evolution of subjectivity.

Given all this, what are the options? If we want to opt for interactionist dualism, would we then hold that little scraps of the extra non-physical stuff enter the story at many successive stages? Or that there might be more or less limited neural communion with a domain of non-physical mind? Alternatively, we might argue that a full-fledged version of experienced mentality comes in at some specific stage. Why would this happen? It would help if there were some *sui generis* features of human brains, or perhaps mammal brains. It is not impossible to argue for such a thing, and many sorts of empirical work might be relevant. Recent work, though, has gone the other way, and has also tended to find apparent marks of some sort of experience in an ever-wider range of animals, however (crabs, bees, octopuses, etc.). The view that looks plausible in the light of recent biology is one that recognizes a gradual evolution of the capacity to experience events. A break between 'higher' and 'lower' animals is looking less and less likely.

---

[18] Popper (1959) discusses the link between simplicity and falsifiability. For the 'Bayesian Occam's razor', see Jefferys and Berger (1992). Some of this literature takes it as constitutive of simplicity that simple theories constrain the data more than complex ones, thereby tightening the link between simplicity and evidence, but that move is quite artificial.

[19] Lycan (2009) also criticizes appeals to parsimony by materialists. I am more sceptical about these arguments than he is. For Lycan, simplicity is a genuine tie-breaker in these sorts of debates when other things are genuinely equal (*ceteris paribus*), but genuine ties are rare and this is not one of them. I agree that ties are rare, but don't think simplicity is a tie-breaker of the kind traditionally envisaged, even in a genuine tie.

If these continuities are embraced, how might we deny materialism? Panpsychism is one route. This is an extreme option, admitted as such by proponents. People see themselves as pushed to it by standard arguments against materialism and dualism. Those arguments against materialism, again, have adequate replies though I am not addressing them here (see note 17). With those arguments defused, panpsychism loses much of its appeal.

Instead of going further down that road, I'll discuss another alternative to materialism, a kind of 'neutral monism'. That name is used for a number of views. What I have in mind is not the 'Russellian' form, in which physics describes pure structure and hence leaves room for intrinsic properties that are 'phenomenal', or close to it. The version I have in mind is seen in Dewey, for example. Dewey (1929). This view holds that what is basic is neither physical nor mental. The patterns and phenomena we associate with those terms ('mental', 'physical') are both manifestations of a more basic reality.

Neutral monism of this kind shades into materialism, on one side, and panpsychism, on the other. It shades into materialism of the more critical kinds, such as what Stoljar calls 'o-physicalism' and the *via negativa* physicalism I mentioned earlier. It shades into panpsychism via a 'panprotopsychism' of the kind Chalmers has defended (Chalmers 2003).[20] That view holds that what is fundamental is not something mental, but something apt to produce the mental, and something other than the merely physical.

I said earlier that the heart of my view, which I see as materialist, is monism and the construction of mind from things that are not mental. Some kinds of neutral monism would seem to pass this test for being materialist. The mental is not derivative on the paradigmatically physical, but it is derivative, in a sense, on the neutral, and hence on something non-mental. (The paradigmatically physical is derivative in the same way.) This blurring of the distinction between materialism and neutral monism might be OK, or we might look for something stronger to mark out materialism, such as the non-derivative nature of the physical.

---

[20] Here is Stoljar's category of an 'o-physical' property:

> a physical property is a property which *either* is the sort of property required by a complete account of the intrinsic nature of paradigmatic physical objects and their constituents *or* else is a property which metaphysically (or logically) supervenes on the sort of property required by a complete account of the intrinsic nature of paradigmatic physical objects and their constituents. (Stoljar 2001, 257)

How does panprotopsychism relate to my fairly broad category of the materialist? The closer the hypothesized quasi-mental entities or properties at the bottom are to being mental, the further the view is from one in which the mental is constructed from the non-mental. The details need to be filled in, and some of the distinctions here are probably not sharp.

In any case, I think there are reasons not to believe a view like this. If I understand the positive content of a neutral monist option of this kind, then something that is central is a kind of extreme anti-reductionism, the claim that it is *not at all true* that the paradigmatically physical side of things is an explainer of many other kinds of goings-on. What we call the 'physical' is just the manifestation of some dispositions, a way things behave in some circumstances, and/or a way visible from some points of view.[21]

But it seems very likely that the physical *is* an explainer—a basis, a 'ground', as people now say—for *some* of the rest, at least. The physical is basic in some respects, not just a set of behaviours in nature on the same level as others. This is evident in the 'easier' cases with respect to the relation between biology and the cognitive side of the mind, especially in the way that the physical and chemical give rise to simpler forms of sensing, memory, and agency.[22]

While neutral monism does make some sense as a rival, it has problems as an overall picture. It denies reasonable kinds of reductionism in its determination to deny a more contentious kind.

Suppose the overall picture had remained one with 'humans versus the rest'—a picture with no gradual evolutionary processes, and humans looking very different from all other creatures. Then I think it would be reasonable to wonder about materialism for longer. More generally, the way to assess materialism is to compare it to other options.

I do accept that 'none of the above' can be a reasonable move in philosophy; then we know we have to think up something new. If materialism was knocked out, we might say that. But materialism is not knocked out. A person can also say:

---

[21] A difference between the 'Russellian' view and the version of neutral monism that I am discussing here is the fact that the Russellian view preserves the basic scientific part-whole hierarchy in most respects. Physics is, in a sense, at the bottom, but physics, together with the edifice built on it, contains a lacuna with respect to the 'phenomenal'. In the view I am sketching (which I take to be close to Dewey's late view, and perhaps also the view in William James (1904) that influenced Russell), the status of the standard scientific hierarchy with respect to part-whole relations and the general direction of explanation is revised in a more wholesale way. The physical processes and phenomena that are usually seen as 'at the bottom', are not really at this 'level', but are just visible in certain contexts—contexts featuring simplicity of interaction and/or contexts involving a particular point of view.

To me, James's own (1904) view seems closer to outright idealism, but if the view is 'neutral', then it has this shape. I don't claim that Dewey consistently endorsed this view, but it is pretty clear in his *Experience and Nature* (Dewey 1929).

[22] Earlier I raised the question of the relationship between my broad form of materialism (monism and the non-mental basis of the mental) and this form of neutral monism. The difference between them can be sharpened if it is claimed (as I would claim) that a complete explanation of the mental has a structure that is in accord with a standard scientific hierarchy with respect to part-whole relations and treats the mental, including the experiential, as a wholly macro-level affair. So, we might recognize a basic view (monism and the non-mental explanation of mind) and a strengthening that take us to materialism as opposed to neutral monism.

'I can't give any detail yet, but something *in this vicinity* is right.' To the extent that the 'vicinity' is filled out, there is an alternative view to consider. Otherwise there is not an alternative to consider—non-materialist mysterianism is not a worked-out rival. And when we look at the views in other vicinities that have been filled out, we find they have worse problems than materialism.

References

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1973) *Belief, Truth and Knowledge*, Cambridge: Cambridge University Press.

Armstrong, D. M. (1989) *A Combinatorial Theory of Possibility*, Cambridge: Cambridge University Press.

Armstrong, D. M. (1993) 'Preface', in *A Materialist Theory of the Mind*, 2nd edition, London: Routledge, xiii–xxiii.

Braddon-Mitchell, David and Frank Jackson (2006) *Philosophy of Mind and Cognition: An Introduction,* 2nd edition, Hoboken, NJ: Wiley-Blackwell.

Chalmers, David J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, New York: Oxford University Press.

Chalmers, David J. (2003) 'Consciousness and its Place in Nature', in *Blackwell Guide to the Philosophy of Mind*, edited by Stephen P. Stich and T. A. Warfield, Oxford: Blackwell, 247–72.

Chisholm, R. M. (1951) 'Reichenbach on Observing and Perceiving', *Philosophical Studies*, 2: 45–8.

Dennett, Daniel C. (1971) 'Intentional Systems', *Journal of Philosophy*, 68: 87–106.

Dennett, Daniel C. (1987) *The Intentional Stance*, Cambridge MA, MIT Press (Bradford Books).

Dewey, John (1929) *Experience and Nature*, revised edition. Reprinted in *John Dewey: The Later Works, 1925–1953,* vol. 1, edited by J.-A. Boydston, Carbondale, IL: Southern Illinois University Press.

Elliot, E. (2015) 'Representation Theorems and the Grounds of Intentionality', PhD thesis, Australian National University.

Geach, Peter T. (1957) *Mental Acts: Their Content and Their Objects*, London: Routledge & Kegan Paul.

Godfrey-Smith, Peter (2016) *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness*, New York: Farrar, Straus and Giroux.

Godfrey-Smith, Peter (2019) 'Evolving Across the Explanatory Gap', *Philosophy, Theory, and Practice in Biology*, 11(1). DOI: https://doi-org.ezproxy.library.sydney.edu.au/10.3998/ptpbio.16039257.0011.001

Hempel, Carl G. (1958) 'The Theoretician's Dilemma', in *Concepts, Theories, and the Mind-Body Problem,* edited by Herbert Feigl, Michael Scriven, and Grover Maxwell, *Minnesota Studies in the Philosophy of Science*, vol. 2, Minneapolis, MN: University of Minnesota Press, 37–98.

Jackson, Frank (1982) 'Epiphenomenal Qualia', *Philosophical Quarterly*, 32: 127–36.

James, William (1904) 'Does 'Consciousness' Exist?', *Journal of Philosophy, Psychology, and Scientific Methods*, 1: 477–91.

Jefferys, William H. and James O. Berger (1992) 'Ockham's Razor and Bayesian Analysis', *American Scientist*, 80(1): 64–72.

Lewis, David (1972) 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy*, 50: 249–58.

Lewis, David (1994) 'Reduction of Mind', in *A Companion to Philosophy of Mind*, edited by Samuel Guttenplan, Oxford: Blackwell, 51–63.

Lycan, William G. (2009) 'Giving Dualism its Due', *Australasian Journal of Philosophy*, 87: 551–63.

Merker, B. (2005) 'The Liabilities of Mobility: A Selection Pressure for the Transition to Consciousness in Animal Evolution', *Consciousness and Cognition*, 14: 89–114.

Montero, Barbara and David Papineau (2005) 'A Defense of the *Via Negativa* Argument for Physicalism', *Analysis*, 65: 233–7.

Place, U. T. (1956) 'Is Consciousness a Brain Process?', *British Journal of Psychology*, 47: 44–50.

Popper, Karl R. (1959) *The Logic of Scientific Discovery*, New York: Basic Books.

Psillos, Stathis (2004) 'Ramsey's Ramsey-Sentences', *Vienna Circle Institute Yearbook*, 12: 67–90.

Ramsey, Frank P. (1931) 'Truth and Probability', in *The Foundations of Mathematics and Other Logical Essays*, edited by R. B. Braithwaite, London: Kegan, Paul, Trench, Trubner and Co, 156–98. (1999 electronic edition.)

Ramsey, Frank P. (1990) 'General Propositions and Causality', in *Philosophical Papers*, edited by D. H. Mellor, Cambridge: Cambridge University Press, 1990, 145–63. First published in 1931.

Ryle, Gilbert (1949) *The Concept of Mind*, London: Hutchinson.

Smart, J. J. C. (1963) *Philosophy and Scientific Realism*, London: Routledge.

Sober, Elliot (1990) 'Let's Razor Ockham's Razor', *Royal Institute of Philosophy Supplement*, 27: 73–93.

Stoljar, Daniel (2001) 'Two Conceptions of the Physical', *Philosophical and Phenomenological Research*, 62: 253–81.

6

Armstrong and Perception

David Rosenthal

1. Armstrong's Causal Account of Perception

David Armstrong's *A Materialist Theory of the Mind* (hereafter *MTM*) was, at its original publication in 1968, the most impressive, thorough, and comprehensive defence of mind-body materialism ever given, and arguably remains so to this day. But it is far more than that. Armstrong's argument for mind-body materialism comes in two stages, first, an exhaustive account of the nature of mental states, and, following that, the considerations that make it overwhelmingly probable that any states that satisfy that account will be bodily, and indeed neural, states. The mental states are neural states and that sustains mind-body materialism.

Armstrong's account of mental states appeals primarily to their causal properties. Mental states are states apt for bringing about particular types of behaviour and, sometimes, apt also for being brought about by particular types of stimuli (*MTM*, 82). To make good on that account requires consideration of all the various types of mental state—a formidable task. Accordingly, well over half of *MTM* is devoted to developing Armstrong's causal account to apply to each of a great variety of mental states, an account that Armstrong argues largely does justice to our common-sense concepts of these states. The result is a remarkably rich and subtle discussion of a comprehensive range of mental-state types, a discussion that repays careful study and in significant measure arguably stands the test of time.

A substantial part of this discussion is devoted to perception and related mental phenomena. Perceptions, Armstrong argues, are cases of acquiring beliefs about the perceptible environment.[1] Beliefs, in turn, are 'mental states apt for selective behaviour towards the environment' (*MTM*, 339) or toward one's own body. Since perceptions are construed as the acquiring of relevant beliefs, they are themselves states 'apt for the bringing about of selective behaviour'. Perceptions are also, Armstrong argues, states 'apt for being brought about by certain

---

[1] Because Armstrong treats bodily sensations as perceptions of bodily conditions (*MTM*, Chapter 14), the beliefs whose acquisition those perceptions consist in will sometimes be about those bodily conditions as well.

David Rosenthal, *Armstrong and Perception* In: *Armstrong's Materialist Theory of Mind.* Edited by: Peter R. Anstey and David Braddon-Mitchell, Oxford University Press. © David Rosenthal 2022. DOI:10.1093/oso/9780192843722.003.0007

[perceptual] situations' (*MTM*, 339, 264), that is, by sensory stimuli, though that aspect of their causal nature figures far less prominently in his discussion.

Being apt for bringing about selective behaviour is a dispositional property; Armstrong characterizes beliefs in dispositional terms. And perceptions, as he notes, are 'definite events' (*MTM*, 214), and hence not simply dispositional. But perceptions are cases of acquiring beliefs, and beliefs are acquired at particular times. So, perceptions, on his account, inherit that time stamp.

Perceiving has traditionally been characterized in qualitative terms. When one sees something red, it's inviting to characterize the relevant perceptual state as itself having some relevant qualitative character. Seeing something red is having a perception *of red*, and a perception's being *of red* is naturally thought of as a matter of the perception itself having some qualitative property.

Armstrong's causal account of perceptions as the acquiring of beliefs avoids any appeal to perceptual states themselves having any such qualitative mental properties. As he notes, the traditional temptation to describe perceptions in qualitative terms is often seen as deriving from a difficulty about describing physical objects as having qualitative properties, such as colours.

Post-Galilean physics requires that the properties of physical objects be susceptible to mathematical treatment. But qualitative properties such as being red—the so-called secondary qualities—seem to resist mathematical treatment. So, some have urged that we in effect relocate the recalcitrant properties to the mind, construing them not as properties of physical objects at all, but of perceptual mental states. Such relocation would leave the physical reality free from any properties that resist mathematical treatment (*MTM*, 271).

A property's being mental does not by itself imply that it is non-physical; without some reason, we cannot conclude that a mental property is automatically non-physical. But because they resist mathematical treatment, physics cannot accommodate the relocated properties, and they would accordingly be non-physical. So, the price of relocating them to the mind would be that the mind would itself be non-physical, at least in that respect. And this would undermine Armstrong's materialist project.[2]

But, as Armstrong argues, there is an alternative to the relocation strategy that handily avoid that anti-materialist conclusion. We can instead identify any properties of physical objects that appear as recalcitrant to mathematical treatment with physical properties plainly susceptible to such treatment. Thus, we can identify colours with physical properties of physical surfaces responsible for the differential reflection of visible light (*MTM*, 272). Physical objects would then have no properties that resist accommodation by physics, and there would be no need to relocate any non-physical properties to the mind.

---

[2] There are many properties physics cannot accommodate that are nonetheless physical, e.g., macroscopic properties such as that of being a tree or a polar icecap. And though such properties do seem in some way reducible to those that physics does accommodate, implementing such a reduction is likely not so easy. But the properties that invite relocation seem plainly to be of a different sort.

The relocation strategy led to the idea that perceptual states have qualitative properties that are also non-physical; so, avoiding relocation leaves Armstrong free to pursue an account of perception independent of any such properties. His account of perceptions as the acquiring of beliefs fits the bill. A perception's being *of red*, for example, is just the acquiring of a belief that something in one's visible environment is red. And such redness is, in turn, a property we can identify with some physically respectable property. Perceptions on this account have no qualitative properties. They would simply represent physical properties that we can understand in non-qualitative terms.

The relocation strategy provides one traditional reason for ascribing qualitative mental properties to perceptual states. But there are others. Several mental phenomena might seem difficult to explain except by appeal to qualitative character. Armstrong singles them out, and argues that we can account for each without any such appeal.

Sensations have often been distinguished from perceptions as states whose only mental properties are qualitative, in contrast with perceptions, which also represent the objects perceived in conceptual terms. Armstrong urges a different way to draw that distinction. Perceptions are acquirings of beliefs, and sensations are a kind of abstraction from perceptions. They are 'the stimulation of the mind by the world in complete abstraction from the further effects of this stimulation on behaviour or impulses towards behaviour' (*MTM*, 254). This fits comfortably within the causal theory of mind, and so avoids any appeal to mental qualities.

Some authors have posited sense impressions as mental states whose distinguishing mental properties are qualitative. Armstrong argues that we can do justice to that notion as well without any appeal to mental qualities. Sense impressions, he urges, are immediate perceptions, perceptions for which the acquiring of beliefs takes place 'without any tincture of inference' (*MTM*, 236). In the visual case, the relevant beliefs will concern only 'a certain group of properties of objects that may be called the *visual* properties' (*MTM*, 234), 'such properties of objects as colour, shape, size, position and motion' (*MTM*, 235). Armstrong acknowledges that disputes are possible about which properties to include for the other sense modalities, but such disputes are unlikely to undermine his reconstruction of the traditional notion of a sense impression.

Mental images are yet another type of mental occurrence traditionally thought to have mental qualities as their distinguishing properties, and raise the most interesting issues. Armstrong acknowledges that mental images 'present an obvious difficulty for a Materialist theory of the mind', but argues that we should construe them as special cases of perceptions that lack two features exhibited by standard perceptions. Mental images are perceptions that, unlike standard cases, do not arise from environmental stimulation, but are instead due to internal causes. And though mental images are otherwise the type of state that could lead one to acquire belief, they fail to do so because we discount the beliefs they would otherwise lead us to acquire. Mental images may in that way 'be said to be "idle" perceptions' (*MTM*, 301).

And, as Armstrong notes, this causal idleness poses a challenge for accommodating mental images within his causal theory of mind. His proposal to meet that challenge relies on resemblance; mental images resemble ordinary perceptions, and the causal theory comfortably accounts for those ordinary perceptions. Mental images are 'events that resemble the acquiring of beliefs about the environment as the result of the action of that environment on the perceiver, although no belief or "potential belief", nor any action of the environment, is involved' (*MTM*, 301).[3]

Though mental images on this account are causally idle, they are states of a type that typically exhibit the distinguishing causal properties of perceptions. A mental image of red resembles an ordinary perception of red because the mental image, like the perception, is the type of state that typically leads to a belief that something red is in front of one. Perceptions consist in the acquiring of beliefs because they do actually lead to those beliefs; mental images are states of the same type, even though they fail to lead to the characteristic beliefs.

And Armstrong urges that '[o]ur introspective awareness of mental images is an awareness of mental occurrences of this sort' (*MTM*, 301). It may be tempting to hold that introspection represents mental images in respect not of their causal properties, but rather of qualitative properties. But, according to Armstrong, there are no qualitative properties of the mental images themselves. The only qualitative properties that figure here are properties of physical objects, such as an object's being red, which the ordinary perceptions that mental images resemble would represent. So one introspects a mental image as being a state of a type that could have led to a belief that a physical object with the relevant physical properties was present. Since that is the mental nature of mental images, that is also how one is introspectively aware of them, as such.

Armstrong's treatment of the three problem cases arguably succeeds in squaring them with his account of perception as the acquiring of beliefs. But that does not address whether an appeal to mental qualitative character might enable a more satisfactory treatment of these cases. I will return to that question at the end of Section 2.

The goal of Armstrong's accounts of sensations, sense impressions, and mental images, and the original account of perception on which those others depend, is to avoid appeal to any mental qualitative character. As he puts it, '[w]e must . . . give an account of perception . . . which does not involve non-physical sensory items'.

---

[3] As Armstrong notes, this echoes his appeal to resemblance in the account earlier in *MTM* of idle wishes and wants, which he argues are simply too weak to cause the characteristic behavioural results (*MTM*, 301).

And he acknowledges that he had 'been unable to see how' to give an account of perception that avoids any such appeal except in terms of the acquiring of beliefs (*MTM*, 217).

If mental qualitative character does, as Armstrong maintains, consist in 'non-physical sensory items', he has no alternative but to avoid any appeal to it. And conceiving of mental qualities as relocated physical properties that seem to resist physicalist treatment does imply that mental qualities are indeed non-physical. If mental qualities are construed as relocated non-physical properties of physical objects, they will carry their non-physical nature with them when relocated to the mind.

The relocation strategy is not the only reason for seeing mental qualitative character as non-physical. It is often held that mental qualities cannot occur without being conscious. If so, consciousness would be our only means of grasping their nature, since first-person access would override any other information we could have about them. Physical properties, by contrast, can always be grasped definitively by third-person access. So, if consciousness is essential to mental qualitative character, mental qualities cannot be physical.[4] Like the adoption of the relocation strategy, the view that consciousness is essential to mental qualities implies that they are non-physical properties.

It is arguable, however, that the relocation story is itself largely responsible for the idea that mental qualities cannot occur without being conscious. Colours, for example, seem recalcitrant to mathematical treatment only insofar as we are aware of them; there is nothing about colours as they occur independently of our awareness that resists mathematical treatment. Physical colour properties seem to call for relocation only as we are conscious of them. So the relocated properties will carry that consciousness along; they will be mental qualitative properties inextricably bound to our consciousness of them (Rosenthal 1999).

But mental qualities need not be conceived of in accordance with the relocation strategy, and one need not take their nature to be accessible solely by way of consciousness. And if one drops those optional assumptions about mental qualities, an account of mental qualitative character is available that fits comfortably within a thoroughgoing materialist view, and indeed within Armstrong's causal theory of mind.

In Section 2, I develop such an account of mental qualities, and in Section 3 I argue that this account avoids all the difficulties Armstrong raises for mental qualitative character. Because the Section 2 account of mental qualities is independent of consciousness, an account is needed to explain how it is that qualitative mental states are sometimes conscious. Section 4 concludes by presenting such an account, one closely related to Armstrong's own account of introspective consciousness, though with some differences I'll comment on.

---

[4] In his earlier *Bodily Sensations* (Armstrong 1962), Armstrong had argued that consciousness cannot be radically wrong about the nature of a particular class of bodily sensations; consciousness would thus largely override any third-person access to sensations in that class. He concluded there that their nature is non-physical (Armstrong 1962, 54–8, 75). In *MTM*, by contrast, he strongly affirms that consciousness can be wrong about mental states of whatever type, perhaps in part because in *MTM* he also rejects any appeal to mental qualitative character.

2. Mental Qualities

Discussions of perception and qualitative mental states generally exhibit a tension between two competing ways of thinking about mental qualities. On one approach, currently very popular, mental qualitative character cannot occur without being conscious; consciousness is built into the very nature of mental qualities. Since, as noted above, first-person access would override any third-person information, consciousness ends up being the only reliable way to learn about mental qualities and their nature. On this 'consciousness-first' approach (Pautz 2013, 195), we learn about mental qualities solely by learning what it's like for one to be in a state of the relevant sort (Chalmers 1996; Nagel 1974; Levine 2000).

This approach makes a mystery of mental qualities. If first-person access dominates our knowledge of mental qualities, we can have no genuinely intersubjective way of picking them out and describing them. First-person access cannot compare your mental qualities with mine; so if we cannot get beyond first-person access, saying that your qualities are like or unlike mine would be empty. It is accordingly this conception of mental qualities that invites thinking that undetectable interpersonal inversion of qualities is possible, or at least conceivable. Only if we can type mental qualities solely by first-person access could one see such an inversion as even conceivable.

But treating consciousness as foundational in this way is by no means the only way to understand the qualitative character of perceptual states. And though Armstrong denies that perceptions exhibit any mental qualitative properties, his remarks about perception fit comfortably with an account of mental qualities that is altogether independent of consciousness.

Perception, according to Armstrong, 'is a matter of acquiring capacities to make systematic physical discriminations within our environment' (*MTM*, 82–3; see 248–50, 255). The beliefs that Armstrong argues are acquired in perceiving enable such discriminative behaviour. It is this tie between perception and discriminative ability that underwrites an account of mental qualities that is independent of consciousness. Rather than see consciousness as basic to our understanding mental qualitative character, we can explain mental qualities by appeal to perceptual discrimination.

An individual's discriminative ability can be determined by very accurate third-person tests. We present the individual with two sample stimuli that are physically the same, and then adjust one stimulus by very tiny increments to make the two samples increasingly different. With colour and sound, we adjust wavelengths, but we can also make incremental adjustments for the chemosensory modalities and for tactile pressure or temperature.

The individual is unlikely to notice that the samples have become slightly different for the first few adjustments; the first adjustment that results in the individual's counting the two as different is a just noticeable difference (JND) for that individual. Such testing can also be done in the opposite direction, making the physical difference between two readily distinguishable stimuli incrementally smaller until the individual perceives them as the same. That so-called matching test has some theoretical advantages over testing for JNDs, discussed below. Still, it will be convenient to cast things mainly in terms of JNDs.

Armed with an exhaustive set of JNDs for all stimuli perceptible by a particular modality, one can construct a quality space that assigns to each stimulus in that modality a relative location. Individuals will, of course, vary slightly in discriminative ability, but one can average to construct a quality space for individuals generally.

Such a quality space represents discriminable physical stimuli. But it points the way to a theory of mental qualities that proceeds independently of consciousness. One can discriminate two stimuli only if one is able to be in two types of perceptual state that differ in ways that correspond to the difference in the stimuli; any discriminable differences between stimuli must reflect differences in the perceptual states one can be in. So the quality space of discriminable stimuli will represent not only discriminable stimuli, but also the various types of perceptual state required to make those discriminations. On the traditional view of mental qualities, moreover, they are the properties of perceptual states that enable such discriminations. Mental qualities differ, on that traditional view, in ways that reflect discriminable differences among perceptible properties. So it is natural to see the quality space as capturing the differences among the mental qualities of a particular modality. We have a quality space that represents discriminable stimulus properties, and a quality space that fixes the mental qualities in virtue of which those properties are discriminable.

JNDs are normally tested by the conscious assessment of whether two stimuli are indistinguishable, but they can also be tested independently of consciousness. One can present the stimuli and enjoin the individual to guess. Indeed, non-conscious perceptual discrimination can be more fine-grained than discrimination that occurs consciously (Scott and Dienes 2010). This quality-space theory of mental qualities delivers an account based solely on perceptual discrimination, with no appeal to consciousness (Rosenthal 1991; 2005, Chapter 7; 2010; 2015).[5]

---

[5] The number of JNDs needed to construct a quality space for all the stimuli discriminable by a perceptual modality would be gigantic. In practice, such quality spaces are constructed using an alternative technique known as multidimensional scaling (MDS). Subjects are asked to estimate the

On quality-space theory, a mental quality of a shade of red is that property of visual states that occupies the same relative location in the quality space of mental colour qualities as the corresponding physical colour shade occupies in the space of physical colour properties. So it is natural to see the mental qualities, thus construed, as not only being qualitative properties, but as also having representational character. Each mental quality represents the perceptible physical property that the mental quality enables one to discriminate. And this account of representation by mental qualities avoids any question-begging appeal to standard conditions and the like.

One might contest whether the properties isolated by quality-space theory are mental qualities. But the only reason to think they aren't is that they can occur independently of being conscious. And it is indisputable that if there are mental qualities, they play the kind of role in perceptual discrimination on which quality-space theory relies. Indeed, that perceptual role is arguably the more revealing of mental qualities than whatever tie they might have with consciousness. The appeal to perceptual role is compelling reason to take the properties quality-space theory describes to be mental qualities.

If one saw mental qualities as having consciousness built in, they could not be representational in the way described a couple of paragraphs back. If consciousness were built in, first-person access to mental qualities would override any third-person information. But we have robust, third-person information about which perceptible properties are represented by which perceptual states. So if consciousness could override any third-person information about mental qualities, they could not be properties that represent corresponding stimulus properties.

Though mental qualities construed in a consciousness-first way resist Armstrong's causal theory and materialist framework, when construed in quality-space terms, mental qualities fit nicely with both. States with different mental qualities enable perceptual discriminations; so, such states will be apt for causing discriminative behaviour. And there is no reason to see mental qualities so construed as non-physical. They are properties of perceptual states posited to explain discriminative ability, not the result of relocating properties of physical objects that appear to resist mathematical treatment.

Nor do mental qualities construed in quality-space terms have any tie to consciousness that might imply non-physical status. Armstrong insists in *MTM* that mental states of every sort can occur without being conscious (*MTM*,

---

relative similarity of strategically chosen stimuli, and statistical methods are applied to the results to generate a quality space. See Clark (1993, appendix), for a useful summary.

MDS is less precise than JNDs for constructing a quality space. But the more important downside is the reliance of MDS on subjective similarity assessments. Because those assessments are subjective, it is unclear how they might be tested subliminally. So the theoretical reliance on MDS likely won't be independent of consciousness. However unworkable a reliance on JNDs is in practice, the theoretical possibility of using JNDs underwrites an account of mental qualities that is independent of consciousness.

113–14), and applies that specifically to perceptions (*MTM*, 231–2). An account of mental qualities that avoids any appeal to consciousness conforms to that sensible insistence.

Armstrong notes that he had been unable to see how to fit perception into a causal theory except on an account of perception as the acquiring of beliefs. Taking perceptions to exhibit mental qualities construed in quality-space terms provides a partial alternative. If perceptions are states with mental qualities, differential mental qualities can discharge the discriminative role played by beliefs in Armstrong's account.

Mental qualities also can discharge some, though not all, of the representational function played on his account by beliefs. Mental qualities represent perceptible stimulus properties; intentional content is needed to represent those stimulus properties as belonging to objects. That second representational function can be discharged by the beliefs that figure in Armstrong's account.

The representational role of mental qualities construed in quality-space terms differs in an important way from the representational role of beliefs. Because the intentional content of beliefs can be described by way of 'that' clauses, it is sentential. Mental qualities, by contrast, represent in a sub-sentential, term-sized way. They do not represent that something is the case; they simply represent types of perceptible stimulus property. A belief might represent that some particular type of object exhibits a particular shade of red; a mental quality would simply represent that shade of red, with no sentential content. A satisfactory account of perceiving must appeal both to mental qualitative character and to intentional content.[6]

An appeal to mental qualities construed along the lines of quality-space theory also allows for more successful treatments of sensations, sense impressions, and mental images than Armstrong's account of perception delivers for these problem cases. Armstrong urges that mental images are states that resemble standard cases of perceptions, but without their causal efficacy. But the respect of resemblance remains unclear on that account, since ordinary perceptions are characterized in respect of causal properties that mental images lack. It is arguably more natural just to say mental images resemble ordinary perceptions in respect of mental quality, which handles the question of respect of resemblance.

---

[6] An account of perception in terms of both qualitative character and intentional content can arguably explain both the low- and high-level contents of perception described in Siegel (2010). Low-level representation proceeds by way of mental qualities and high level by way of intentional content.

A convincing theory of perception that relies on both types of mental property is found in Sellars (1956). Armstrong would rightly object that Sellars' account of mental qualities relies on a version of the relocation story, with its attendant difficulties vividly evident in Sellars (1965). But that aspect of Sellars' view can be detached from the rest, leaving something very close to quality-space theory; see Rosenthal (2016).

Armstrong argues that we are introspectively aware of mental images in respect of their resembling perceptions with standard causal properties. But introspection arguably never represents mental states in respect of their causal properties. As with other causal ties, we must learn from experience how introspected states connect causally with other mental states and with behaviour and stimulus conditions. Introspection does not provide that information. So, an account of mental images on which they resemble ordinary perceptions in respect of qualitative character fits more comfortably with what we independently take introspection to reveal.

Armstrong sees sense impressions as perceptions for which the beliefs acquired in the visual case pertain only to 'a certain group of properties of objects that may be called the visual properties' (*MTM*, 234), 'such properties of objects as colour, shape, size, position and motion' (*MTM*, 235). Given the exclusive focus on such perceptible properties, it is more natural to construe sense impressions as states whose only mental properties are mental qualities. Mental qualities, moreover, have no intentional content, so that purely qualitative states cannot figure as steps in inferences. So taking mental qualities to be the only mental properties of sense impressions fits well with Armstrong's observation that sense impressions occur 'without any tincture of inference' (*MTM*, 236).

Similar comments apply to Armstrong's construal of sensations as 'the stimulation of the mind by the world in complete abstraction from the further effects of this stimulation on behaviour or impulses towards behaviour' (*MTM*, 254). It is arguably more natural to see sensations as states whose sole mental properties are mental qualities. The only downside of taking these problem cases to exhibit mental qualities would be if that undermined Armstrong's materialist program. The quality-space account avoids that difficulty.

3. Armstrong's Challenges to Mental Qualities

Some of Armstrong's reasons to reject any appeal to mental qualities rest on the assumption that they would be 'non-physical sensory items' (*MTM*, 217). And on some accounts, especially any that rests on the relocation story, mental qualities are indeed non-physical. But nothing about quality-space theory implies that.

But Armstrong raises other challenges as well. One is that countenancing mental qualities runs the risk that what we have knowledge of is in the first instance just those mental qualities, and we can get knowledge of physical objects only by inferring from those mental qualities (*MTM*, 217–18). Mental qualities would constitute a veil of ideas, separating us epistemologically from physical reality.

But we can accommodate mental qualities without seeing them as the primary objects of empirical knowledge. Mental qualities construed in quality-space terms have a representational aspect; a mental quality of red represents the physical colour. They are mental properties by means of which we have knowledge of physical reality. In that way, they are like beliefs themselves. If I have a belief that there is a red object in front of me, nobody would suggest that my having knowledge about that red object requires that I infer from the occurrence of that belief. Since the belief represents the object, the belief can by itself provide me with knowledge about the object. The same holds for mental qualities. Their representing the perceptible stimulus properties of objects can provide one with knowledge about those properties. We don't infer from our being in states with mental qualities any more than we infer from our having beliefs.

Armstrong also rejects a construal of mental qualities on which they are causally idle, 'mere phenomenological facts, mere accompaniments to our acquiring beliefs about the current state of our body and its physical environment' (*MTM*, 218). And he rightly rejects mental qualities so construed. There would be no reason to countenance mental qualities construed as causally idle. But mental qualities on quality-space theory are by no means causally idle. They are mental properties that enable discrimination among perceptible properties of physical objects and events.

A more telling objection stems from the possibility of three physical items very close in colour, so that the first is indistinguishable from the second and the second from the third, though the first and third are distinguishable. The difference from first to second and from second to third does not permit differentiation, but the cumulative difference from first to third does. This connects with the very JND apparatus that underlies quality-space theory.

Since one cannot discriminate the first from the second, Armstrong assumes that the mental qualities corresponding to those two would be identical. '[T]he two sensory items...that we have when we look at the [first] two pieces *actually are identical in colour*' (*MTM*, 218, emphasis Armstrong's), and similarly for mental qualities corresponding to the second and third. If the identity holds for the first and second mental qualities and for the second and third, it would have also to hold for the first and third. But the first and third physical samples are distinguishable. So the mental qualities corresponding to them cannot be identical; identical mental qualities would imply indistinguishability of physical samples. Transitivity can fail for the distinguishability of physical samples, since that is not a property of those samples. But transitivity cannot fail for identity of mental qualities themselves.

The obvious move, Armstrong notes, is to deny that our introspective knowledge of the mental qualities is infallible.[7] If introspective access to one's own mental qualities need not always reflect their nature, mental qualities could be introspectively indistinguishable but not on that account identical. In the case at hand, since the first and third mental qualities are plainly distinct, at least one of the other two pairs must be as well, despite being introspectively indistinguishable.

---

[7] He speaks of 'incorrigible knowledge' (*MTM*, 219), but what it is at issue is not immunity from being corrected, but not be able to be wrong.

But Armstrong argues this move won't work. The phenomenological appearances when stimuli are indistinguishable would seem clear, he urges, and the only reason to think they might be erroneous is that, if they are accurate, there is apparent trouble for the appeal to mental qualities. Better then to jettison that appeal. Moreover, he argues, 'those who support the analysis of perception as involving sensory items, regularly allege that . . . their view alone . . . does phenomenological justice to the perceptual facts' (*MTM*, 219). If mental qualities are invoked to do phenomenological justice to our perceptual states and it turns out that they cannot, so much the worse for mental qualities.

But that argument makes it clear that the mental qualities Armstrong here rejects are those conceived on a consciousness-based model, on which first-person access is decisive about their mental nature. If first-person access is not the last word about their nature, our first-person appearances could diverge from that nature. Only on a consciousness-based conception of mental qualities could they be invoked to do phenomenological justice to our perceptual states. Conceived in quality-space terms, by contrast, first-person access to mental qualities need not match their mental nature. This argument against mental qualities goes against only a consciousness-based conception.

Although Armstrong's argument from transitivity leaves a quality-space conception of mental qualities untouched, transitivity considerations do affect the appeal to JNDs in constructing quality spaces. Being just noticeably different is not transitive; a first stimulus may be just noticeably different from a second and the second from a third even though the first is indistinguishable from the third. So JNDs do not provide a reliable way to individuate stimulus properties, and by extrapolation they cannot reliably individuate the corresponding mental qualities.

A workaround is available, however, by relying not on JNDs, but on matching. Matching is also intransitive; a first stimulus may be indistinguishable from a second and the second from a third even though the first and third are distinguishable. Pairs of mental qualities for the first and second stimuli and for the second and third would match, though mental qualities corresponding to the first and third stimuli would not. But adapting a suggestion of Goodman's (1966, 272), we can count two mental qualities as identical in type just in case there are no mental qualities matched by one but not the other.[8]

Armstrong offers yet another challenge to mental qualities, in somewhat the same spirit as the argument from transitivity. This argument rests on the apparent indeterminacy of perception. One may see a speckled hen, it seems, without its being determinate how many speckles one sees. One may see one object as being larger than another without its being determinate how much larger it is. If mental qualities are responsible for such indeterminate perceptual appearances, their nature must be indeterminate as well.

---

[8] Goodman (1966, 311–13) also notes difficulties pertaining to symmetry and proposes an elaborate solution along somewhat similar lines (1966, 313–17).

Again, Armstrong considers and dismisses an account on which the mental qualities are determinate, but we are not introspectively aware of them as determinate in the relevant respects. His objection is the same; though mental qualities are 'specially postulated to do phenomenological justice to perception [, they] are now credited with characteristics that lie quite outside perceptual awareness' (*MTM*, 220–1). But again the argument is in effect not against mental qualities, but only against conceiving of them in a consciousness-based way. Mental qualities construed on a quality-space conception are unaffected by the objection.

Armstrong has another argument, however, made almost in passing. If mental qualities are determinate but our awareness of them reflects a phenomenological indeterminacy, why aren't the mental qualities redundant? 'Why not simply postulate the speckled physical surfaces and indeterminate awareness (perception) of those surfaces?' (*MTM*, 221). Given that our awareness of the physical objects is indeterminate, why countenance determinate, intermediate mental properties?

This challenge points nicely to Armstrong's account of perception in terms of the acquiring of beliefs, which he goes on to develop and defend a few paragraphs later. Why not simply rely on relevant beliefs for the indeterminate awareness of determinate physical perceptible properties? Beliefs seem tailor-made to represent indeterminately.

But the determinate way in which mental qualities represent does not make them redundant when construed along quality-space lines. Mental qualities are the mental culmination of the perceptual process, and reflect the determinate character of that process. They may often be somewhat less determinate than the corresponding perceptible properties of perceived objects, but there is no reason to expect mental qualities to reflect the degree of indeterminacy that occurs subjectively in the phenomenological appearances.

In this respect, the mental qualities posited by quality-space theory have an advantage over the beliefs posited on Armstrong's account of perception. The beliefs that figure in Armstrong's account do reflect the phenomenological indeterminacy perception. But as noted in Section 2, perceptual discriminating that is not conscious can be more fine-grained than conscious discrimination (Scott and Dienes 2010); judgments based on unconscious perceptual information can be more accurate than judgments based on conscious perceptual information.

So there must be mental properties of perception that reflect the greater accuracy that unconscious perception can have compared with conscious perception. To apply this to the indeterminacy Armstrong appeals to, it may be that an unconscious perception of a speckled hen is less indeterminate in respect of the number of speckles than a conscious perception of the hen. We need mental properties of perceptual states that can underwrite and explain such experimental findings.

Armstrong's beliefs cannot fit the bill. But the mental qualities posited by quality-space theory, which occur independently of consciousness, can. They are the mental properties that reflect whatever perceptual determinacy may be evident in such experimental findings. Mental qualities construed in quality-space terms are not redundant.

4. Qualitative Consciousness

Armstrong's challenges to mental qualities all rest on construing mental qualities in consciousness-based terms, as mental properties whose nature we know about only from first-person access. This reflects to some extent his treatment of mental qualities in the earlier *Bodily Sensations*. His departure in *MTM* from that earlier work was to jettison mental qualities altogether, still construed in consciousness-based terms. But one could instead simply reject the consciousness-based view, and adopt an approach based on perceptual role, such as quality-space theory.

And that strategy fits with Armstrong's view in *MTM* about mental states generally. We can never, he maintains there, rely on the accuracy of first-person access (*MTM*, 100–3; 326). We always need sound third-person backing to determine the nature of mental states of whatever sort. That is how quality-space theory treats mental qualities.

The replies of Section 3 to Armstrong's objections to mental qualities all rely on being able to distinguish mental qualities from our subjective awareness of them. That distinction is basic to quality-space theory, and indeed any perceptual-role approach, since fixing mental qualities by perceptual role need make no appeal to consciousness.

But even though we can account for mental qualities in a way that doesn't require states with mental qualities to be conscious, such states are sometimes conscious. And since quality-space theory doesn't explain what it is for qualitative states to be conscious, we need a separate account to explain that. Armstrong's own account of what it is for mental states of any sort to be conscious works very well for that purpose.

Mental states are not in themselves conscious; being conscious is not a property that mental states of any sort automatically have. Rather, a state's being conscious is an add-on, something distinct from the state itself. '[C]onsciousness is no more than *awareness* (perception) of inner mental states by the person whose states they are' (*MTM*, 94; emphasis Armstrong's). A state's being conscious simply consists in one's being aware of that state in a suitable way.

Armstrong approvingly notes Kant's appeal to inner sense, which takes 'sense perception as the model for introspection' (*MTM*, 95; cf. 323ff.).[9] Absent such higher-order perception, no mental state is conscious; when one is aware of a mental state by way of such higher-order perception, that state is conscious. As with mental states of every type, such higher-order perceptions would not be conscious on their own. A higher-order perception is conscious only if there is a yet higher-order perception about the first higher-order perception.

This account is compelling. One of the few things we can be confident of in connection with consciousness is that if one is in some mental state but wholly unaware of being in it, that state fails to be conscious. By contraposition, a conscious state is a state that one is in some suitable way aware of being in. 'Consciousness . . . is simply awareness of our own state of mind' (*MTM*, 95). The only question, then, is how one is aware of a state when that state is conscious. On Armstrong's view, we are aware of it by perceiving it.

Armstrong's view of consciousness as a kind of internal perception relies on his view of perception as the acquiring of beliefs. It is the intentional content of a belief acquired about a mental state one is in that is responsible, on his account, for one's being aware of that state when it is conscious. One is aware of the state in virtue of the belief acquired in having a higher-order perception, a belief that represents one as being in that state.

Mental qualities construed in quality-space terms also represent things, and might also do so on a consciousness-based view. But perception as Armstrong conceives of it involves no mental qualities. Intentional content is the only type of mental representation that can on his view explain an awareness of a mental state.

Since it is the intentional content of an acquired belief in virtue of which one is aware of one's conscious states, we can ask why that intentional content should occur in connection with perceiving. Why wouldn't an intentional state to the effect that one is in the target state serve the purpose just as well even if it occurred independently of anything one might count as perceiving? Why isn't the perceptual context for the relevant belief idle?

Elsewhere (Rosenthal 2005; 2012; 2018), I have argued that the best explanation of how one is aware of a mental state when that state is conscious is that one has a higher-order thought (HOT) to the effect that one is in that state. This invokes the aspect of Armstrong's account that makes one aware of the target mental state, leaving behind the perceptual context.

If one is in a mental state but reports not being in that state, we regard that state as not being conscious. Reporting that one is in a mental state is the standard indicator, both in everyday and experimental contexts, that a state is conscious. Reporting, moreover, consists in expressing a state whose intentional content is the same as the report. So, if one can report being in a mental state, the best explanation of that ability is that one is in an intentional state whose content is that one is in that state (Rosenthal 2005, Chapter 2). Intentional content is crucial in explaining how we are aware of a mental state when it is conscious.

---

[9] The term 'introspection' is often restricted to the attentive awareness of a mental state that one is deliberately focusing on, but Armstrong uses the term to refer both to mental states being conscious in the ordinary, inattentive way and to the attentive, focused awareness of one's mental states (see *MTM*, 91).

Does the positing of a higher-order perception, construed in Armstrong's way, have any advantage over the more modest positing of a HOT? One advantage might be that perceiving is subjectively unmediated; nothing seems to mediate our perceptions of external objects. If there is any mediation, mental or otherwise, we are subjectively unaware of it. And the same holds for our first-person awareness of our conscious states; it seems subjectively unmediated. That may make perception inviting as a model for introspective awareness. But the same holds for our thoughts; we are sometimes unaware of any mediation between a thought and what it is about. And we can stipulate that this is so with HOTs. Subjective immediacy cannot provide a reason to prefer higher-order perception.

But there is another reason for Armstrong's preference of a perceptual model. Perception monitors our surroundings, registering the objects and events that might affect us. And Armstrong sees our higher-order awareness as a kind of scanning process (*MTM*, 94, 102–3, 107, 163, 324), in effect, monitoring our mental states.

Perceptual monitoring of our external environment, moreover, has great utility. And Armstrong argues that the same holds for the internal scanning of our mental states. Such scanning, he urges, has great utility, since it is required for purposive behaviour and problem solving. In thinking purposively or solving a problem, we go through a succession of mental steps. And to go from one step to the next, Armstrong maintains, requires at each stage that one be aware of the last mental step, so that feedback from one's awareness of each previous mental step can inform how to proceed with the next (*MTM*, 163).

This is equally so, he urges, for deciding on a course of action (*MTM*, 140) and performing a calculation. '[I]f our mind is to work purposively... we must have awareness of our minds' (*MTM*, 163). Only by being aware of one's 'current mental state... can we adjust mental behaviour to mental circumstances... Only if we do become so aware we will we know what to do [i.e., what to think] next' (*MTM*, 327). These claims fit well with Armstrong's scanning metaphor for introspective awareness. A state's being conscious, on that model, is due to the mind's scanning itself and determining mentally that the state occurs.

There are, of course, multistep mental processes in which feedback from earlier steps is needed to determine which subsequent mental steps to take. Armstrong urges that such feedback must be due to awareness of the earlier steps. But that is by no means obvious. Each state in a mental process will affect subsequent mental steps because of that state's mental properties, independently of whether the state is conscious. Each step in the process will cause subsequent steps in virtue of the earlier step's mental properties. The causal powers of those mental properties do not depend on the state's being conscious.

Armstrong argues that 'if we are doing a calculation "in our head" we will need to become aware of the current stage in the mental calculation that we have reached. Only if we do become so aware will we know what to do next' (*MTM*, 327). But each step in the calculation consists in one's being in an intentional state with content that bears on that calculation. That intentional content should be enough to determine the next step. And if it is not, there is no reason to think that an awareness of being in a state with that content would help. The higher-order awareness adds no useful information beyond the content of the first-order state itself. Armstrong is right that 'without information of some sort about the current state of our mind, purposive trains of mental activity would be impossible' (*MTM*, 327). The content of the first-order state by itself provides that information.

These considerations can be generalized. The utility of being in any mental state is due to that state's mental properties, especially its representational properties, such as intentional content. It is unclear what utility might be added by that state's being conscious. The utility of a state is likely due solely to its representational mental properties and not at all, or if at all only marginally, to its being conscious (Rosenthal 2008; 2012, §V; 2019).[10]

There are several reasons this point is easy to miss. We are aware of a state's utility only when that state is conscious; when the state isn't conscious, one is unaware of being in the state, and so unaware of any utility it may have. Also, a huge number of the mental states that figure in our daily lives are conscious; so it's tempting to restrict attention just to them. And there is a well-entrenched tendency to think about things in terms of their utility.

Armstrong notes that multistep purposive thinking sometimes seems to happen during sleep; one may go to sleep puzzled by a problem and awaken with an elaborate, multistep solution. He suggests that some subliminal introspective awareness of the various steps may have occurred in such cases, though he also expresses uncertainty about that proposal (*MTM*, 164). And there is good reason for doubt. Subliminal perception is simply perception that isn't conscious; so subliminal introspective awareness, by analogy, is simply introspective awareness of which one is not in turn aware. And that is simply the standard case; when we are in everyday situations aware of our mental states, we are not also aware of being aware. Armstrong plainly has in mind something else by subliminal introspection, but it is unclear what might be.

---

[10] It is important that the argument against significant utility of a state's being conscious does not rest on a higher-order theory of consciousness. The argument appeals simply to consideration of which mental properties are likely responsible for such utility. Still, the argument does fit comfortably with a higher-order theory, and in that way provides some indirect support for it.

It is better just to accept problem solving during sleep as evidence that introspective awareness is not needed for purposive, multistep, mental processes. And if introspective awareness does not confer utility in the way Armstrong suggests, the main reason for construing introspective awareness on a perceptual model is substantially weakened. Positing a perceptual context for the intentional content in virtue of which one is aware of one's conscious states is arguably idle. We should instead adopt a HOT theory of such awareness.

The issue between a perceptual model, construed on Armstrong's theory of perception, and the HOT theory is minor compared to the difference between higher-order theories generally and first-order views, which reject any appeal to higher-order awareness in explaining consciousness. First-order views typically adopt a consciousness-based view of mental states, and see consciousness as an intrinsic property of mental states.[11]

Such first-order approaches have difficulty accommodating mental states that fail to be conscious, for example, states that occur in masked priming (Marcel 1983; Breitmeyer and Öğmen 2006) or in blindsight (Weiskrantz 1997; 2009). If being conscious is intrinsic to mental states, why do any occur without being conscious? And first-order views typically take a mental state's being conscious to be an irreducible monadic property of that state, giving no informative account of the nature of that property, urging that we can learn about consciousness only by what consciousness itself tells us. All that makes a mystery of consciousness, encouraging the view that no informative explanation is possible. We are better accepting a higher-order theory, one very much along the lines of Armstrong's in *MTM*.

References

Armstrong, D. M. (1962) *Bodily Sensations*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.

---

[11] An exception is the global neuronal workspace theory, according to which a state is conscious if its content is globally available for psychological processing (Dehaene and Naccache 2001), independent of any higher-order awareness. But it is unclear why global availability should result in a state's being conscious. Indeed, there seem to be counterexamples in both directions; some unconscious beliefs and desires likely have global effects, and peripheral vision even when conscious rarely does.

One type of higher-order theory construes the higher-order content as internal to each conscious state, thereby reflecting a consciousness-based view (e.g., Kriegel 2009). Armstrong argues against such a view; 'it is impossible that the introspecting and the thing introspected should be one and the same mental state' (*MTM*, 324). Armstrong appeals here to his scanning metaphor; 'the operation of scanning and the situation scanned must be "distinct existences"' (*MTM*, 107). For a compelling, independent argument against Kriegel's account of higher-order content as intrinsic to the conscious state, see Phillips (2014).

Breitmeyer, Bruno G. and Haluk Öğmen (2006) *Visual Masking: Time Slices through Conscious and Unconscious Vision*, 2nd edition, Oxford: Oxford University Press.

Chalmers, David J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, New York: Oxford University Press.

Clark, Austen (1993) *Sensory Qualities*, Oxford: Clarendon Press.

Dehaene, Stanislas, and Lionel Naccache (2001) 'Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework', *Cognition*, 79(1–2): 1–37.

Goodman, Nelson (1966) *The Structure of Appearance*, 2nd edition, Indianapolis, IN: Bobbs-Merrill.

Kriegel, Uriah (2009) *Subjective Consciousness: A Self-Representational Theory*, Oxford: Oxford University Press.

Levine, Joseph (2000) *Purple Haze: The Puzzle of Consciousness*, New York: Oxford University Press.

Marcel, Anthony J. (1983) 'Conscious and Unconscious Perception: An Approach to the Relations between Phenomenal Experience and Perceptual Processes', *Cognitive Psychology*, 15(2): 238–300.

Nagel, Thomas (1974) 'What Is It Like to Be a Bat?', *The Philosophical Review*, 83(4): 435–50.

Pautz, Adam (2013) 'Does Phenomenology Ground Mental Content?', in *Phenomenal Intentionality: New Essays*, edited by Uriah Kriegel, Oxford: Oxford University Press, 194–234.

Phillips, Ben (2014) 'Indirect Representation and the Self-Representational Theory of Consciousness', *Philosophical Studies*, 167(2): 273–90.

Rosenthal, David (1991) 'The Independence of Consciousness and Sensory Quality', *Philosophical Issues: Consciousness*, 1: 15–36. Reprinted in Rosenthal (2005), 149–74.

Rosenthal, David (1999) 'Sensory Quality and the Relocation Story', *Philosophical Topics*, 26(1–2): 321–50. Reprinted in Rosenthal (2005), 135–48.

Rosenthal, David (2005) *Consciousness and Mind*, Oxford: Clarendon Press.

Rosenthal, David (2008) 'Consciousness and Its Function, *Neuropsychologia*, 46(3): 829–40.

Rosenthal, David (2010) 'How to Think about Mental Qualities', *Philosophical Issues: Philosophy of Mind*, 20: 368–93.

Rosenthal, David (2012) 'Higher-Order Awareness, Misrepresentation, and Function', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 1594 (May 19): 1424–38.

Rosenthal, David (2015) 'Quality Spaces and Sensory Modalities', in *The Nature of Phenomenal Qualities: Sense, Perception, and Consciousness,* edited by Paul Coates and Sam Coleman, Oxford: Oxford University Press, 33–65.

Rosenthal, David (2016) 'Quality Spaces, Relocation, and Grain', in *Sellars and His Legacy*, edited by James R. O'Shea, Oxford: Oxford University Press, 149–85.

Rosenthal, David (2018) 'Seeming to Seem', in *The Philosophy of Daniel Dennett*, edited by Bryce Huebner, New York: Oxford University Press, 133–64.

Rosenthal, David (2019) 'Consciousness and Confidence', *Neuropsychologia*, 128: 255–65.

Scott, Ryan B. and Zoltán Dienes (2010) 'Knowledge Applied to New Domains: The Unconscious Succeeds Where the Conscious Fails', *Consciousness and Cognition*, 19(1): 391–8.

Sellars, Wilfrid (1956) 'Empiricism and the Philosophy of Mind', in *Minnesota Studies in the Philosophy of Science*, vol. 1, edited by Herbert Feigl and Michael Scriven, Minneapolis, MN: University of Minnesota Press, 253–329.

Sellars, Wilfrid (1965) 'The Identity Approach to the Mind-Body Problem', *The Review of Metaphysics*, 18(3): 430–51.

Siegel, Susanna (2010) *The Contents of Visual Experience*, New York: Oxford University Press.

Weiskrantz, Lawrence (1997) *Consciousness Lost and Found: A Neuropsychological Exploration*, Oxford: Oxford University Press.

Weiskrantz, Lawrence (2009) *Blindsight: A Case Study Spanning 35 Years and New Developments*, 2nd edition, Oxford: Oxford University Press.

7

Does *A Materialist Theory of the Mind* Give Us a Reply to the Knowledge Argument?

Frank Jackson

1. Introduction

My thesis is a simple one. If the accounts David Armstrong[1] gives in *A Materialist Theory of the Mind* of perception and of colour are correct, there is a relatively straightforward reply to the knowledge argument. Well, not quite. At various points, I will tweak what Armstrong says (at one point it is a bit more than a tweak). I will take the liberty of suggesting various improvements to the accounts he gives of perception and of colour, and will proceed to argue that the improved versions allow a relatively straightforward reply to the knowledge argument. A side benefit of our discussion of colour is that no fault disagreements about something's colour are possible, consistent with an objectivist–realist position on colour. (I will talk interchangeably of realist and objectivist views, meaning views that hold an object's colour is independent from what it causes by way of visual experiences; colour is, in this regard, like, for example, shape. Sometimes 'realist' is reserved for the doctrine that objects sometimes possess the objective properties in question. This is by the way here, as I will presume that some objects are coloured.)

2. Perception and the Collection of Information

The ways in which physical objects are affected by their surroundings are major sources of information. That's why weather balloons, space probes, petrol gauges, and so on are so useful. For example, the orientation of the pointer in a gauge on the dashboard of my car is affected by how things are in my car's petrol tank. This gives me information about how much petrol is left in the tank. In this example, there is an important distinction between the nature of that which delivers the information and the information delivered. The pointer's being orientated thus

---

[1] My debts are too many to list, but a special mention of David Armstrong is appropriate. I had the great good fortune to be taught by him in my first year as a philosophy undergraduate.

Frank Jackson, *Does* A Materialist Theory of the Mind *Give Us a Reply to the Knowledge Argument?* In: *Armstrong's Materialist Theory of Mind.* Edited by: Peter R. Anstey and David Braddon-Mitchell, Oxford University Press.
Frank Jackson 2022. DOI: 10.1093/oso/9780192843722.003.0008

and so does not *in itself* say that the amount of petrol remaining in the tank is so and so, which is why we need to learn to read petrol gauges. *Mutatis mutandis* for the other examples listed. A bit of jargon: I will say that in cases like these, the connection is *extrinsic*. The states that carry the information do not, by their very nature, ‘say’ what information they carry.

Our bodies are physical objects that are affected by their surroundings in ways that provide information about those surroundings. When someone enters a building with wet hair, you learn that their head has been in contact with water recently—most likely it is raining outside. But some effects on our bodies, and more particularly our brains, are special. Perceptual experiences are informationally rich, causal responses to how things are that are special in that the connection between an experience and some (not all) of the information it carries is *intrinsic* in the following sense: although a perceptual experience is distinct from that which it carries information about, the nature of the experience *qua* experience determines the information carried. When something looks to be moving, the information provided by the experience is part and parcel of the experience’s being the kind of experience it is. Again, when something feels warm or to have sharp edges, the information provided by the experience comes from the nature of the experience *qua* experience. A nice question is exactly how to specify the kind of information carried by the nature of a perceptual experience *qua* kind of experience it is. We will steer clear of this question by focusing the discussion to come on relatively non-controversial candidates (for the case of visual experience, these candidates are listed in Section 3). This is one strand in the thought of those who argue that perceptual experiences represent that things are thus and so, where this is understood as a thesis about their sensory nature.[2] Intrinsicality explains the fact that although young children and dogs don’t get information about how much petrol remains in the tank from a petrol gauge, or about atmospheric pressure from instruments carried by weather balloons, they do get information about, for example, motion, temperature, and shape from having experiences as of things looking to be moving, feeling hot or square, and so on. Getting information from having perceptual experiences does not require lessons—lessons that young children and dogs have not had—in the way that getting information from petrol gauges and weather balloons does.

Two caveats. Sometimes the information isn’t information in the strict sense. It is, I take it, always putative information—it is always the case that a perceptual experience points by its very nature to things being a certain way—however, when the vertical lines in the Heller Illusion look curved, what the experience ‘says’ is that the lines are curved, whereas in reality they are straight. I will follow the not uncommon practice of using ‘information’ in the more inclusive sense, as does

---

[2] See, for example, Tye (1995). In later work, Armstrong talks in terms of representation—‘perceptions are representations’ (Armstrong 1999, 131).

Armstrong (see, e.g., *MTM*, Chapter 10, §I). Second, the information is not always about how things are around the perceiver—when you feel water going down your throat, you are have a perceptual experience that carries information about how things are in your throat.

What's the best explanation of intrinsicality, of the fact that information carried by perceptual experiences is intrinsic to them? Armstrong favours the simplest explanation (to put matters in our terms, he does not talk in terms of intrinsicality). He holds that perception is the acquisition of belief (*MTM*, Chapter 10, §I). The reason the connection between something's looking to be square and the information that it is square is intrinsic is that the perceptual experience is nothing more than the acquisition of the belief that the thing is square.[3]

Simplicity is good, but maybe in this case it is too good to be true. Among the questions to be asked of the view that perception is nothing more than the acquisition of belief, two stand out. First, what about cases where belief is not acquired, cases where subjects believe, rightly or wrongly, that they are under an illusion or are hallucinating and, as a result, do not acquire the belief in question? Second, how can a view of this kind handle the distinctive phenomenology of perceptual experience? Although philosophers are divided over the extent to which belief and the acquisition of belief have a phenomenology, it is widely held that if they do, it is very different from the phenomenology of perceptual experience.

Armstrong is well aware that there are cases of perception without belief. He insists that in such cases there is at least an inclination to believe, or maybe there is a suppressed belief—the perceiver acquires a state that would be the relevant belief except for . . . , where the dots are filled in with whatever it is that convinces the perceiver that they are under an illusion or are hallucinating (see *MTM*, Chapter 10, §IV). But consider the Heller Illusion mentioned earlier. For experienced customers—those familiar with these kinds of examples of perceptual illusions—there is no inclination whatever to believe that the lines are curved. And even those who are unfamiliar with the example, quickly lose any inclination to believe that the lines are curved once they hold a straight edge up to them. Are subjects like these, nevertheless, in a state that *would* be the belief that the lines are curved except for . . . ?

As is well known, counterfactual analyses have many problems. I will take these as read and press an objection of principle. The objection is that the lines' looking curved is a categorical feature of the perceptual experience someone has when undergoing the Heller Illusion. It is a feature of the state they are in fact in.

---

[3] Maybe, we should talk of information that *something or other* is square, not of some given thing being square. However, the question of what it takes to get information about some *given* object raises questions orthogonal to our interests here. We will allow ourselves to fudge.

I think Armstrong should have gone down a different but nearby path, one faithful to his thought that we should approach perceptual experience in a way that makes its informational role central, but which construes its nature as residing in a categorical feature. Imagine that Mary is undergoing the Heller Illusion but knows full well that she is under an illusion. There is a way the lines would have to be to be in accord with her perceptual experience, and she will know what that way is. She will know that the lines would need to be curved. If that were not true, Mary would not know what makes the Heller Illusion count as an illusion, and of course she does. Or consider the following kind of situation, loosely based on what is depicted in the film *Total Recall.* I undergo a series of perceptual episodes, each lasting, let us say, a day, and ending in a short period of unconsciousness. Each episode is coherent within itself, and each suggests that its immediate predecessor was an unusually vivid and realistic dream from which I have just awakened. After a while, I become totally confused. I have no inclination, suppressed or otherwise, to believe that things are a certain way. Nevertheless, what will be true is that, for each episode, there is a way things would have to be to be in accord with my experiences during that episode—the experiences I have point to a certain way things are, and in that sense urge on me an account of how things are—and I know what that way is. That's how I know that each is in conflict with its immediate predecessor, and is the key as to why I end up so confused.

My suggestion is that Armstrong should have said that perception is the acquisition of a state that urges that things are a certain way, and the nature of a perceptual experience *qua* perceptual experience lies in how the state is urging things to be. Another way of putting essentially the same idea is to say that perceptual experiences are acquisitions of continually renewed invitations to believe that things are thus and so.[4] Of course, the urgings I am talking about typically lead to belief, and it is very plausible that the explanation of how animals evolved to have these urgings will lie in the survival value of the beliefs they typically lead to.

You now have before you my first tweak: in order to meet the challenge of perception without belief, replace Armstrong's account of perception as the acquisition of belief by one in terms of the acquisition of states urging that things are thus and so. But what about the other question mentioned earlier, the challenge of explaining the distinctive phenomenology of perceptual experience? Can our tweaked version of Armstrong's account of perception do this? I think it can. The key is to note what's special about the urgings we are talking about, and here we are following Armstrong's lead in the following sense: when he argues that perception is nothing but the acquisition of belief, he emphasizes that the belief in question is special in various ways (*MTM*, Chapter 10, §IX).

---

[4] The term 'invitation' comes from Foster (2000, 98) but his overall position on perception is very different from the Armstrongian one sketched here. Foster is an idealist.

We will now review what's special about those urgings. Our focus will be on visual perception, as that will help us when we come to discuss colour and what to say about Mary in the black and white room—to put matters in terms of the version of the knowledge argument in Jackson (1982).

3. What It Takes to Be As a Visual Experience Urges Things To Be

Earlier, we said that the information a perceptual experience carries is intrinsic in the following sense: the kind of experience it is determines the information it carries—where given the remarks of Section 2, the information it carries is to be understood as the view about how things are being urged upon the subject in virtue of having the perceptual experience. This implies that its experiential nature determines how things have to be in order to be in accord with the experience, because how things have to be is the information carried. Or, to say it in terms of representation: a perceptual experience, by virtue of the kind of experience it is, represents that things are a certain way, and that way—its content—will be what we call above, how things are being urged to be by the experience. But what is that way and how is it special? (What follows has been noted by philosophers but is in fact a bit of folk theory. Anyone with normal vision will be able to check on what I am about to say.)

It is special in three ways. First, typically it is very rich. How things are being urged to be, the content, concerns shape, size, colour (in the inclusive sense, not just the hue sense), and various spatial relationships to the body of the subject having the experience—motion relative to that body, distance away from it, orientation relative to that body (tilted with respect to the perceiver's line of sight, in front and to the left, etc.), and so on. What is more, this rich picture of how things are is being urged to obtain here and now; that is, at the time of having the experience. Second, there aren't separate experiences for each item of information. There isn't, for instance, one experience that urges that things are a certain way regarding colour, and another that urges that things are a certain way concerning shape or motion. There is a single experience that represents all of shape, colour, motion, size, orientation, etc. What's more, colour is represented as the colour of something with a shape that is or is not moving; distance away is represented as the distance away of something that is or is not moving and is of a certain shape, colour, and size; and so on. In short, typically, what gets represented—in the sense of how things are being urged to be—is that there are objects, each having all of colour, shape, a certain orientation with respect to the perceiver's body, and so on. Finally, the objects having these various properties are represented as standing in various relationships to each other as well as to the perceiver. In sum, what's being urged upon subjects having visual perceptual experiences is a whole way that the space before them is filled out with objects having colours, shapes, and all that, and standing thus and so with respect to each other and the subject (as Peacocke (1992) captures nicely with the term 'positioned scenarios'.) All this has an important implication for what it takes to be as a typical visual experience says things are; what it takes, that is, for the information that is intrinsic to a typical visual experience to be information properly speaking. What it takes is for there to be a range of objects arranged thus and so at the time in question, in front of the perceiver with properties like shape, colour, and orientation.[5]

Now we can ask the 'phenomenology' question. What would it be like for a subject to be in a state that delivered the information *intrinsically* that this is how things are; that continually urged, by virtue of the kind of experience it is, that this is how things are? I submit that it is very plausible that it would seem to the subject like being in a kind of direct acquaintance relationship with objects that had in fact the properties that things look to the subject to have. The sense datum theory was a mistake but one grounded in an insight, the insight commonly referred to in terms of the diaphanousness or transparency of perceptual experience.[6] In some sense, things looking thus and so is a matter of things being thus and so—that's the insight behind diaphanousness—but what needs to be thus and so is how things are being urged to be, not how some things (sense data) actually are. Sense data were posited in order to get the phenomenology right. If what we have just said is correct, we account for the phenomenology without the implausible metaphysics.

We can now explain the appeal of transparency or diaphanousness. An experience's nature is given by what it says about how things are; that's why describing a visual experience seems to be nothing more than describing how the things we see are, despite the fact that, when under illusion, it isn't true that we are describing how things are: the lines in the Heller Illusion aren't curved. We can also explain what happens when subjects report on the nature of various illusions. In order to find out whether or not the vertical lines in the Heller Illusion look curved—something you need to do in order to ascertain that it is indeed an example of an illusion—you have to look at the diagram itself. How so? The lines are not curved.

---

[5] Typically in front of the perceiver, but drivers see objects in rear vision mirrors *as* being behind them, despite the fact that they are looking forwards.

[6] See Moore (1922); Harman (1990). It is noteworthy that Harman says, 'I will say little that is original and will for the most part merely elaborate points made many years ago..., points that I do not think have been properly appreciated' (Harman 1990, 33); he gives special mention to *MTM*. I find myself in a somewhat similar position. When I became convinced we should reject sense data and move to a broadly Armstrongian view, I thought the kind of reply I give here to the knowledge argument was sort of obvious. I have been surprised at the degree of resistance by some I expected to be allies, as evinced in responses to Jackson (2003), one of which, (Byrne 2006), included the term 'jaw-dropping'.

The vertical lines' looking curved is a fact about you—more especially, about what the diagram does to you. All the same, you have to focus on the diagram to check on how the lines look to be. It would seem that you are focusing on the wrong place! The solution to the puzzle is that how the diagram looks to you to be is a matter of how the perceptual experience caused by the diagram urges the *diagram* to be; and, in order to report on this, you need to focus on the diagram. Here it is important to bear in mind the distinction between how your experience urges things to be and the extent to which you are aware of how your experience urges things to be. To *report* on how the experience caused by the diagram urges things to be, you need the second, and that's what you get when you attend carefully to the diagram.[7]

We are now in a position to discuss what Armstrong says about colour; this is where the major tweak occurs.

4. Colour

Armstrong supports physical realism about colour; he was one of the first to expound it. Physical realism about colour has two parts. One part holds that colours are properties of objects that are every bit as objective as, for example, shapes. If we concentrate on surface colour, the view is that colours are properties of objects' surfaces, properties whose nature is not in any way constituted by our responses to those surfaces. This is the realism cum objectivism part of the doctrine. The other part holds that these properties are physical properties in the sense of being the sorts of properties optical physics talks about. A popular suggestion (e.g., Byrne and Hilbert 2004) is that they are certain kinds of spectral reflectance profiles of objects' surfaces. Under the influence of Hilbert (1987), Armstrong (1999, 125) came to accept this version of physical realism, having earlier favoured the view that colours are certain micro-physical properties of surfaces. This popular version permits a simple explanation of colour constancy. Despite big changes in the light reflected from the surfaces of objects as the day progresses, the colours objects look to have often remain more or less constant. The explanation is that the reflectance profiles remain the same through the day. Our discussion will assume this version of physical realism; nothing hangs on this particular choice.

There are two serious problems with physical realism. The first relates to the special informational role of perception we have already talked about. Perceptual experiences represent—that is, as we are spelling this thought out, they urge—that there are coloured, shaped, moving or not moving, at so and so a distance away, etc. objects, and they do so by virtue of being the kind of experiences that they are. This is how the information they provide gets to be intrinsic to the experiences, which is what makes these experiences so useful, and, we argued, explains the phenomenology of perception. But no one thinks that carrying information about spectral reflectances is intrinsic to visual perceptual experiences. We can acquire information about motion and shape simply by seeing objects as approaching us or as round; we cannot learn about spectral reflectances simply by seeing objects as coloured. Learning about spectral reflectances took serious science. The upshot is that realists about colour cannot be *physical* realists about colour. Spectral reflectances aren't the kinds of properties of objects that visual perceptual experiences give information about by virtue of being the kind of experiences that they are. The same goes for any plausible alternative candidates to identify with the colours that physical realists might put forward, which is why nothing hangs on our having chosen to frame the discussion in terms of spectral reflectances.

---

[7] Armstrong holds that phenomenal consciousness is a matter of having a mental state directed towards a mental state (Armstrong 1968, Chapter 6, §IX). If what we say above is correct, this isn't quite right. Consciousness, or at least consciousness of phenomenal nature, is a matter of having a mental state directed towards the relevant *content*. What one is conscious of is not the state but what the state says about how things are.

The second problem concerns the evidential role of colour appearances. Our reason for holding that some object is, say, green is typically that it looks green in certain circumstances, and likewise for the other colours. Now physical realists cannot explain this fact—and it needs explanation, it isn't a brute fact—in terms of an object's being green being, in one way or another, a matter of its looking green in those certain circumstances: that's what the 'other' party, the subjectivists cum response dependence theorists about colour, say. Nor can they explain this fact by making it a central tenet of physical realism that the property of being green is the physical property—a certain spectral reflectance, as we are supposing—that causes the object to look green in those circumstances. To do that would mean that what *makes it the case* that the physical property in question is identical with being green is what it does to perceivers.[8] (Subjectivists in ethics don't turn themselves into objectivists by holding that goodness is the property that causes (or actually causes) certain feelings of approval rather than the feelings themselves, and then arguing that science tells us that what causes these feelings of approval is one or another objective property.)

What realists about colour must do, it seems to me, in order to explain how objects' looking green can provide good reason for holding them to be green is to draw on the core idea behind the informational approach to perception. They must view colour properties as intrinsic to the nature of visual perceptual experiences: the very nature of the experiences informs us about the relevant properties—about, that is, colours. But, of course, reflectances are not properties intrinsic to the nature of visual experiences, as we have lately noted. At the end of Section 5, we explain how realists (but not of course physical realists) about colour can allow that the very nature of visual experiences informs us about colour.

---

[8] For more on this and why what I say about colour in Jackson (1998, Chapter 4) is mistaken, see Jackson (2019b).

The upshot is that realists about colour must look elsewhere for the properties to identify the colours with. I think that Armstrong tells us where to look. That's the business of the next section.

5. Armstrong on the Concept of Colour

Armstrong urges that the concept of colour is that of properties unknown to perceivers that stand in various relations.[9] His thought is that when an object looks to be a certain (surface) colour to you, you are in a state that gives information (in the intrinsic sense) about how the surface in question is related to other surfaces in terms of likenesses (or similarities) and unlikenesses (or differences), without giving information about the properties of the surfaces that constitute the likenesses and unlikenesses. If the spectral reflectance explanation of colour constancy is correct, plausibly the likenesses and unlikenesses will be in how alike or unalike the reflectances are, but that's not something given by how the objects look. Face recognition gives us a model for thinking about his thought. When you see someone as someone you have seen before, that person's face is seen as being like a face seen in the past, but the respect in which the face seen now and the face seen in the past are alike is not given in perception (which is why it takes work to find out how the brain 'does' it). The survival value of face recognition is obvious, and the same is true of the ability to spot similarities and differences in objects' surfaces, whether or not you can spot the respect in which they are similar and different.[10]

Which surface similarities and differences are revealed when objects look coloured? An attractive answer is: those captured by the colour solid, which represents degrees of likeness and unlikeness with regard to hue, saturation, and brightness/lightness. They plausibly correspond to the information that comes from having visual experiences, for the fact that the colour solid captures the relations in question is something available simply by having good vision.[11] But what's important for our purposes is the core idea: the concept of colour is the concept of occupying a certain place in a network of similarities and differences.

Now for the major tweak. Take Armstrong's account of the concept of colour and treat it as an account of colour properties. Armstrong holds that such and such a colour is identical with the property that occupies so and so a place in the network we have just been talking about. Instead he should, I suggest, hold that such and such a colour is having the property that occupies so and so a place in that network. Here's a way of thinking of what I am recommending. Take the question, What is it to be the author of *David Copperfield*? Two answers suggest themselves: (1) It is to be Charles Dickens; (2) It is to be whoever wrote *David Copperfield*. The second answer is the one that provides the model I am suggesting for colour properties.

---

[9] (*MTM*, Chapter 12, §II); see also the discussion in Armstrong (1999, 128–9).

[10] I take this paragraph to expound what Armstrong has in mind at a number of places in *MTM* (Chapter 12, §II).

[11] It is, accordingly, no surprise that many have been attracted to this idea, in one form or another. See Hazen (1999) and references therein for the relevance of the idea to the semantics of colour terms. For objections to views of this kind, see Boghossian and Velleman (1991), around p. 94; for a reply, see, for example, Jackson (2019b).

This allows Armstrong to hold onto his realism–objectivism about colour. An object's having a surface property that stands in such and such a location in a network is a property its surface possesses independently of perceivers (and will obtain in virtue of facts about spectral reflectances of surfaces or some such): perception reveals the property to us in the sense that seeing something in good lighting conditions causes us to go into a state that urges that the property is possessed, without in any way creating the property. And it avoids the problem we raised for physical realism about colour. For it is plausible that the property in question is intrinsic to visual experiences. We *see* the similarities and differences and we *see* their extent: the surfaces of objects that look red look to be alike, and look to be markedly different from the surfaces of objects that look green; the surfaces of objects that look pink look to be somewhat like the surfaces of objects that look red and somewhat like the surfaces of objects that look white; and so on. As far as I can see, there is no reason why Armstrong could not accept a view of this kind. (And, of course, we could call this a kind of physical realism about colour if we expanded what counts as a physical property so as to include locations in similarity and difference networks that supervene on physical nature in the other sense of 'physical'.)

This tells us why looking green in the right conditions is good evidence for being green, the question we noted earlier to be an urgent one for objectivists about colour. Because having a certain location in a network of similarities and differences is intrinsic to the experience of something's looking green, the experience singles out that very property as the one the experience urges an object to have. The objectivist does not need to posit as a brute, unexplained fact that looking green can be good evidence for being green. Its being good evidence follows from the very nature of the experience plus a plausible view about how the nature of an experience lies in how it urges its subject to take things to be.

Now for the short digression on no fault disagreements about colour.

6. How No Fault Disagreements About Colour Are Possible

On the view outlined above, colours are locations in a similarity and difference network, locations occupied, as a matter of fact, by certain reflectance profiles, and when people form judgements about an object's colour on the basis of how it looks, they are forming judgements about the location of the object's surface in that network.

This means that no fault disagreements about colour are possible. Reflectance profiles are complex. There are many ways they can be alike and can be different, each equally objective. In this regard, they are no different from, for example, economies, cities, and films (movies). When our visual systems classify objects as alike or different, what they are doing is latching onto one or another aspect of these complex profiles, and there is no particular reason why the visual systems of different subjects might not do this somewhat differently. Consider the well-known case of subjects with equally good vision in general, and colour vision in particular, who differ over which Munsell chip is true blue—is blue with no hint of any other hue.[12] It is possible that one perceiver is right and the other is wrong. Anyone sympathetic to realism–objectivism about colour must allow for this possibility. But it would be nice if we supporters of objectivism about colour had another option. The symmetry of the situation suggests that it would be arbitrary to insist that we must always plump for one subject getting it right and the other getting it wrong. On the view outlined in Section 5, we have another option. The two perceivers may have visual systems that differ slightly in where they locate the reflectance profiles in question in the network definitive of being blue, and both may be right. They may each be latching onto perfectly objective features of the profiles, but they are latching onto different, perfectly objective features.[13]

7. A Reply to the Knowledge Argument

What follows is a short version of things I have laboured elsewhere, see, for example, Jackson (2019a).

Mary, while in the black and white room, does not know what it is like to see something as green; that's the intuition. But not knowing what something is like is ambiguous.

First, it may mean not knowing the properties of whatever it is. In that sense, the reply to the knowledge argument based on what we have said to date is a flat out denial. She does know the properties of something's looking green. It's a state that, by its very nature, urges on a subject that things are thus and so, or, equivalently, the information being urged to the effect that things are thus and so is intrinsic to the experience. And what's the thus and so? That also is something she knows: it is that the surface of the object occupies a certain location in a network of similarities and differences. And Mary can know that. Mary, given her vast physical knowledge, is in a position to record the spectral reflectances of the surfaces of objects outside the black and white room and note the extent to which they are alike and different. Once she's done that, she can find the location of each surface's reflectance profile in the network of similarities and differences definitive of the colours. She will then have, as a result of complex calculations, the same information about surfaces that those of us with good colour vision get when we see things in good lighting conditions. (Of course, she will have, in addition, information about the properties of the surfaces that are responsible for their being at those locations.) Here's a helpful analogy. Armed with a grammar book, someone who cannot speak or write English can work out which sentences in English are or are not grammatical. But we competent speakers and writers of English do not need a grammar book. Some sentences simply look or sound grammatical to us, and some look or sound the opposite. In seeing or hearing them as grammatical, we see or hear them as alike in a certain way, a way which is hard to describe but easy (for us) to spot. But, of course, we competent writers and speakers of English do not acquire information denied to the user of the grammar book. We aren't accessing properties—being and not being grammatical—hidden from users of the grammar book. We simply have a different and easier way of accessing the very same information and properties.

---

[12] There are a number of papers on true blue; see, for example, Byrne and Hilbert (2007) and references therein.

[13] For a fuller discussion and related issues, see Jackson (2019b).

Of course, Mary's perceptual experiences will lack the phenomenology of having something looking to be, say, green. As we saw earlier (Section 3), that phenomenology is plausibly explained by the highly distinctive way that certain perceptual experiences urge on those who have them views about how things are, and, as far as colour is concerned, where the surfaces of seen objects stand in a certain network of similarities and differences. Her experiences lack this useful property, at least while she is inside that room, and that's why they have a different 'feel' from ours.

There is, however, a sense—a different one, thus the talk of ambiguity above—in which she does not know what it is like to see things as coloured. This sense is akin to that in which the face-blind do not know what it is like to recognize faces. Because she has not had experiences that carry the similarity and difference information about surfaces intrinsically, she cannot recall them (the experiences), reflect on them, recognize them, and so on. This is because some abilities require the right kind of causal history, and the ones we are talking about fall into that category. That's the sense in which Mary will not know what it is like to see something as, for example, green, much as Lewis (1990) and Nemirow (1980) say.

What should Mary be thinking before she is due to be released? That she will soon have experiences that urge on her views about similarities and differences between the surfaces of the objects she sees (and maybe other relations the surfaces stand in), and do so intrinsically in the sense that having the experiences will be enough. No longer will she need to carry out elaborate calculations. In that sense, she will know what to expect. And she will look forward to acquiring an ability to recognize, think about and recall these experiences, for those abilities await her actually having the experiences.

References

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul (2nd edition 1993).

Armstrong, D. M. (1999) *The Mind-Body Problem*, Oxford: Westview Press.

Boghossian, Paul and David Velleman (1991) 'Physicalist Theories of Color', *The Philosophical Review*, 100(1): 67–106.

Byrne, Alex (2006) 'Review of *There's Something about Mary*, edited by Peter Ludlow et. al.', Cambridge, MA: MIT Press, 2004', *Notre Dame Philosophical Reviews*, 1: 2006.01.20.

Byrne, Alex and David R. Hilbert (2004) 'Hardin, Tye, and Color Physicalism', *Journal of Philosophy*, 101(1): 37–43.

Byrne, Alex and David R. Hilbert (2007) 'Truest Blue', *Analysis*, 67(1): 87–92.

Foster, John (2000) *The Nature of Perception*, Oxford: Oxford University Press.

Harman, Gilbert (1990) 'The Intrinsic Quality of Experience', in *Philosophical Perspectives*, 4, *Action Theory and Philosophy of Mind*, edited by James Tomberlin, Atascadero, CA: Ridgeview Press, 31–52.

Hazen, A. P. (1999) 'On Naming the Colours', *Australasian Journal of Philosophy*, 77(2): 224–31.

Hilbert, David R. (1987) *Color and Color Perception: A Study in Anthropocentric Realism*, Stanford, CA: CSLI.

Jackson, Frank (1982) 'Epiphenomenal Qualia', *Philosophical Quarterly*, 32(127): 127–36.

Jackson, Frank (1998) *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford: Clarendon Press.

Jackson, Frank (2003) 'Mind and Illusion', in *Minds and Persons*, edited by Anthony O'Hear, Royal Institute of Philosophy Supplement: 53, Cambridge: Cambridge University Press, 251–71.

Jackson, Frank (2019a) 'The Knowledge Argument Meets Representationalism about Colour Experience', in *The Knowledge Argument,* edited by Sam Coleman, Cambridge: Cambridge University Press, 102–17.

Jackson, Frank (2019b) 'How to Be an Objectivist about Colour', *Phenomenology and the Cognitive Sciences*, special issue, edited by Tony Chen and Paul Snowdon, 18: 819–31.

Lewis, David (1990) 'What Experience Teaches', in *Mind and Cognition*, edited by William G. Lycan, Oxford: Blackwell, 29–57.

Moore, G. E. (1922) 'The Refutation of Idealism', in *Philosophical Studies*, London: Routledge & Kegan Paul, 1–30.

Nemirow, Laurence (1980) 'Review of Thomas Nagel, *Mortal Questions*', *The Philosophical Review*, 89: 475–6.

Peacocke, Christopher (1992) *A Study of Concepts*, Cambridge, MA: MIT Press.

Tye, Michael (1995) *Ten Problems of Consciousness*, Cambridge, MA: MIT Press.

8

Closing (or at Least Narrowing) the Explanatory Gap

Katalin Farkas

1. Introduction

Rereading the Preface to the 1993 paperback edition of David Armstrong's *A Materialist Theory of the Mind* (*MTM*, originally published in 1968) brings back vividly the personality of the author for those, who, like me, knew him. Armstrong had a strong feeling that he was an important part of the history of philosophy.[1] When he describes his theory in relation to those of other important figures, such as J. J. C. Smart and U. T. Place, this reads like a twentieth-century account of the culmination of the development of the World Spirit (except that nothing is further from the author's intention than declaring that the world is moved by a mind!). The book is indeed one of the finest expressions of a relentless materialist (or physicalist—I'll use these terms interchangeably) vision of the mind, and Armstrong's causal theory of the mind has proved to be a powerful framework to explain various mental phenomena. But philosophers who accept some version of materialism still often think that their theory leaves something unexplained about the nature of conscious experiences. This contrasts with theories about physical phenomena, which are not considered to leave something unexplained in the same way. This phenomenon is known as the 'explanatory gap'.

In this chapter, I revisit the issue of the explanatory gap. In Section 2, I will recall a version of this idea from an influential paper which originated the very term: Joseph Levine's (1983) paper 'Materialism and qualia: and the explanatory gap'. In Sections 3 and 4, I will relate the issue to Armstrong's discussion of secondary qualities and bodily sensations, and to Locke's idea that only the ideas of primary qualities resemble the quality that causes them. In Section 5, I argue that the specific explanatory question Levine asks is this: 'Why this phenomenal character, rather than any other, is attached to this physiological process?'.

---

[1] If an anecdote is permitted here: in a conversation, Armstrong once expressed scepticism about the advantages of blind refereeing for journals. Someone objected: 'But, David, if you submitted a paper to a journal and it contained a mistake, you'd prefer if that mistake was identified by a referee, and the paper wasn't published just because you wrote it?' Armstrong replied: 'If I make a mistake, it is an event of philosophical significance.' This was, of course, largely a joke—but not entirely, I think.

Katalin Farkas, *Closing (or at Least Narrowing) the Explanatory Gap* In: *Armstrong's Materialist Theory of Mind.*
Edited by: Peter R. Anstey and David Braddon-Mitchell, Oxford University Press. © Katalin Farkas 2022.
DOI: 10.1093/oso/9780192843722.003.0009

In Sections 6 and 7, I argue that this question can be answered. First, we should look for a fit not between the realizer of a role and the phenomenal character, but between the role itself and the character. Further, there is a natural fit between the phenomenal character of experiences and their functional roles: for example, pains feel inherently unpleasant, and that explains why they cause avoidance behaviour. In Section 8, I sketch a possible background theory that explains this fit, which I call Phenomenal Functionalism: a view that is meant to be similar to Phenomenal Intentionality and holds that the functional role of an experience is grounded in its phenomenal features. In Section 9, I discuss some other possible gaps in our understanding the relationship between the mental and the physical and conclude that the fit between functional role and phenomenal character goes a long way, though probably not the whole way, to closing the explanatory gap.

2. The Explanatory Gap

As Armstrong explains in the Preface, the causal theory, his own materialist account of the mind, has two steps. First, we give an analysis of mental features, based on the insight that a mental feature is apt to cause certain ranges of behaviour, and it is apt for being produced by certain ranges of stimuli. Second, we claim that what is apt for causing and being caused this way is a physical state of the brain. That physical state is then identified with the mental state (Armstrong 1993, xiv). These were originally regarded as contingent identities, but subsequently, Saul Kripke (1972) influentially argued that identity statements were necessary, even when empirically discoverable. Reflecting on this development, Armstrong notes that Kripke's argument works only for rigid designators, whereas he intends the physical description of these states as non-rigid designators. Hence he thinks that Kripke's argument against materialism fails (Armstrong 1993, xiv).

Kripke's argument against materialism has indeed left many philosophers unconvinced, but a subsequent reflection on the identity statements in the focus of Kripke's inquiry, by Joseph Levine, proved to be very influential. In this section, I will recall Levine's well-known argument for an 'explanatory gap' between the physical and mental description of certain phenomena.

Levine opens his paper (Levine 1983) by recalling Kripke's argument against mental-physical identities, such as pain is C-fibre firing. Following Kripke (1972), Levine compares the following identity statements:

(1) Pain is C-fibre firing.
(2) Heat is the motion of molecules.

As is well known, Kripke argues that the first is false, the second is true. Levine has some doubts about Kripke's argument; he thinks that the metaphysical thesis of the distinctness of mind and body is not conclusively proved by Kripke. However, he wants to make a different point in his paper. He claims that even if Kripke were wrong, and contrary to what he says, both statements were true, there would still be a significant difference between them. The difference is not metaphysical, but rather epistemic: it relates to the explanatory power of the statements. The second is fully explanatory, Levine claims, but the first leaves something crucial unexplained.

In setting up the problem, Levine considers a causal-functionalist account of heat and pain, which consists of the two steps of Armstrong's causal theory, as sketched in the first paragraph of this section (though Levine doesn't particularly mention Armstrong). First, we identify the causal or functional role of pain and heat. Pain typically 'warns us of damage, it causes us to attempt to avoid situations we believe will result in it, etc.' (Levine 1983, 357). Heat is 'responsible for the expansion and contraction of mercury in thermometers, causes some gases to rise and others to sink, etc.' (Levine 1983, 357). Second, we look for the state that occupies this causal role and we identify these states with the original phenomena: it's the motion of molecules in the case of heat, and C-fibre firing in the case of pain.

According to Levine, the resulting statement (2), the identification of heat with the motion of molecule, expresses an identity that is *fully explanatory*. First, prior to the discovery of identity, the causal role of heat exhausts our notion of it: what we understand by the phenomenon of heat is fully accounted for by the causal role it plays. Second, knowledge of chemistry and physics makes it intelligible how the motion of molecules plays the causal role associated with heat. So, once we have identified the causal role, and discovered what plays this causal role, everything is explained.

The case of pain and C-fibre firing is different, Levine claims. What is similar is this. The causal role of pain is at least part of our concept of pain. And in the case of pain too, after we identify the causal role of pain, we discover a mechanism, C-fibre firing, that plays this causal role. But there is a crucial difference: our notion of pain is *not exhausted* by the causal role. The causal role is part of the notion, but there is more, namely, the phenomenal or qualitative character of pain. And while our knowledge of physiology makes it intelligible how C-fibre firing plays the causal role associated with pain, there is nothing in our knowledge of physiology that explains why C-fibre firing gives rise to the phenomenal properties of pain. As Levine puts it: '. . . there seems to be nothing about C-fiber firing which makes it naturally "fit" the phenomenal properties of pain, any more than it would fit some other set of phenomenal properties' (Levine 1983, 357).

According to Levine, the same story goes, for example, for the sensations of colours. Sensations of red and green are apt to be caused by certain stimuli, which we may call physical red and physical green (first step). Now consider the receptors and physical processes that are responsible for responding to physical red and physical green, and call these physical stories R and G (second step). R is supposed to give rise to, or correlate with, the phenomenal feel of red experiences, and G the phenomenal feel of green experiences. But there is no explanation for this in R or G: 'R doesn't really explain why I have one kind of qualitative experience . . . and not the other' (Levine 1983, 358). Support for this claim is provided by imagining an inversion between the two qualitative characters, while the physical stories remain the same. The inversion seems perfectly conceivable, and this underlines the point that the phenomenal feel of red has no more intelligible connection to R than the phenomenal feel of green.

Talking of colour inversion may bring Locke's name to mind, and indeed, Levine says that the same point was captured by Locke who thought that 'two sets of phenomena—corpuscular process and simple ideas—are stuck together in an arbitrary manner' (Levine 1983, 359). Levine is right to trace back this idea to Locke, and he may have also traced it further back to Descartes. Neither Descartes nor Locke supported an identity theory, but both dealt with the question of the relationship between the ideas of secondary (and tertiary) qualities, and the physical and physiological processes that precede or cause the formation of these ideas. Both Descartes and Locke claim that *it is simply ordered by God which conscious experience goes with which physical process*. Invoking God's will in this way means, for both of them, that there is no intelligible explanation of the connection—that is, an explanation that is intelligible for us.

3. Secondary and Tertiary Qualities

Levine's (1983) paper was published in the period between the first 1968 edition of *A Materialist Theory of the Mind* and the paperback edition (Armstrong 1993), but it isn't included among the philosophical developments that Armstrong chose to reflect on in the Preface. Nor there is a mention of the explanatory gap in Armstrong's (1999) *The Mind-Body Problem*. However, it is instructive to recall Armstrong's discussion of secondary qualities and bodily sensations and compare it to the issue of the explanatory gap.

Both pain and colour sensations are analysed by Armstrong as perceptual experiences (Chapters 12 and 14 of *MTM*) and are given a materialist account within his general materialist theory of perception as an acquisition of beliefs. The particular problems that need to be handled are the problems emerging from treating these experiences as perceptions.

In particular, if colour experiences are perceptions, then colours have to be analysed as perceivable physical properties of physical objects. There are well-known objections to this view, but Armstrong thinks the objections can be overcome, and colours can be identified with micro-physical properties of objects. Bodily sensations are perceptions of various parts of the body. Pain, for example, is a perception of certain disturbances in the body, most probably of the stimulation of pain receptors.

The peculiarity of these experiences, according to Armstrong, is that they don't offer any clue about the nature of the physical or physiological properties that they are experiences of:

> [W]hile we recognize by sight that all red things have something in common, sight does not inform us what that common property is. In the same way, we recognize by bodily perception that the class of felt disturbances called 'bodily pains' all have something in common. But bodily perception does not inform us what that common feature is. (*MTM*, 314)

> [W]e recognize by sight that red things differ among themselves in respect of redness. There are different shades of red. In the same way, certain pains resemble each other and differ from other pains. We recognize different sorts of pain. In the case of colours, however, we need not concede that vision informs us of the nature of the differences involved in being different shades of the same colour. In the same way, we need not concede that bodily perceptions inform us of the nature of the difference of bodily disturbance involved in the different sorts of pain. We are simply informed that the disturbances do differ in some respect. (*MTM*, 314–15)

Perhaps this is the point where something like the explanatory gap appears in Armstrong's theory. The issue is not the same as the one discussed by Levine. Armstrong's remarks concern the relationship between the felt quality of the experience and the property that's perceived in having that experience, while Levine is interested in the relationship between the experience and the realizing brain (or nervous system) state. Still, the issue is a certain lack of intelligible connection. Viewed from the side of the experiences, there is nothing that tells us about the nature of the physical property that is perceived in the experience. So viewed from the other direction, the nature of the physical property will probably not explain why the experience feels the way it does.

Note the similarity between Armstrong's remarks and, again, Locke's ideas on primary and secondary qualities. Locke thought that the ideas of primary qualities *resemble* those qualities that cause them; whereas the ideas attached to what we call 'secondary qualities' don't resemble their causes, because those causes are in fact combinations of primary qualities of insensible parts: 'the ideas of primary qualities of bodies, are resemblances of them, and their patterns do really exist in the bodies themselves; but the ideas, produced in us by these secondary qualities, have no resemblance of them at all. There is nothing like our ideas, existing in the bodies themselves' (Locke 1975, II. viii. 15).

Ideas of sensations like pain, sometimes called ideas of 'tertiary qualities' are treated by Locke in the same manner as ideas of secondary qualities. Tertiary qualities are bare powers of objects; that is, a combination of the primary qualities of insensible parts, that cause certain ideas in us, but again, the ideas don't resemble the causes. The difference between ideas of secondary and tertiary qualities is that, in the latter case, we are not even tempted to think that their causes resemble them, while in the former case, we are so tempted.

4. An Idea Can Be Like Nothing But an Idea

As many have pointed out, the notion that ideas of primary qualities resemble the qualities that cause them is puzzling. Something has a circular shape if it takes up a certain space within a boundary whose points are equidistant from a central point. How could an idea resemble this, when ideas are either immaterial—as Locke thought—or realized in a brain state that is certainly not circular itself? As Berkeley aptly remarked, 'an idea can be like nothing but an idea' (1949, 44). Ideas of primary qualities will no more resemble the qualities that cause them than ideas of secondary qualities.

The suggestion that ideas could represent things in the world by resembling them is certainly discredited in contemporary theorizing about representations, and yet the Lockean idea survives in various forms. One, it seems to me, is Armstrong's view above: when we recognize that all red things have something in common, the idea does not inform us of the nature of the shared property—that is, that it's some microphysical property of surfaces.

The situation is supposed to be different for primary qualities. But is it? Locke's list of primary qualities is solidity, extension, figure (shape), motion or rest, and number. Consider the primary qualities involved in my current experience. I see *one stationary* coffee cup on the table, with a *continuous smooth* surface and a *cylindrical* shape. As a matter of fact, these qualities are realized by a combination of the primary qualities of insensible particles: a structure of not one, but *many* particles, with *space between them*, and these particles may *move* while the cup is stationary. Where is the resemblance? And we could hardly discover the microphysical realization of these perceivable properties just from the ideas of the properties (perceived shape, size, and motion) themselves. So primary qualities seem to have the same status in this respect as secondary qualities. The point, as it is sometimes remarked, is the difference between the manifest and the scientific image, rather than between primary and secondary qualities.

Interestingly, elsewhere Armstrong acknowledges what seems to me a similar point. He introduces the topic of secondary qualities with the well-known observation that the scientific image of the world, at least since Galileo, provides no place for the secondary qualities, since physical theories do not attribute such qualities to the fundamental constituents of the physical world (like particles or fields). He adds that some philosophers hold that therefore secondary qualities must qualify the mind, rather than the physical world. Armstrong responds by endorsing an objection made by Berkeley: in a visual experience, he says, for example, colour and visible extension 'are inextricably bound up with each other' (*MTM*, 272). So if we relegate colour to the mind, we should do the same with visible extension.

All these points seem to be manifestations of the same underlying issue. Our awareness of primary qualities in perceptual experiences surely contributes to the phenomenal character of these experiences alongside our awareness of secondary and tertiary qualities. Insofar as these are all ideas, that is, (properties of) experiences, they have the same ontological status. And even if they are all material, none of them will 'resemble' their extra-experiential causes more than any other: a perception of a blue circle will be no more circular than it is blue. Further, neither of them will inform us, just by the way they appear in our awareness, about their microphysical realization.

When Locke says that for secondary qualities, God assigned simple ideas to corpuscular processes, but could have chosen to assign different ones, he is motivated by the thought that the ideas of secondary qualities do not resemble their causes. One of the places that Levine references in Locke's *Essay* (Locke 1975, II. viii. 13) explicitly mentions the lack of 'similitude' between these ideas and the movement of insensible particles affecting our senses which cause these ideas. Although Levine says that Locke captures the same point that he (that is, Levine) is trying to make, surely he doesn't mean the view that only primary qualities resemble their causes. So we should see if Levine is more successful than Locke and Armstrong in drawing a contrast between primary and secondary qualities along these lines.

5. Why This, Rather Than That?

We have seen in Section 2 that Levine thinks that there is an explanatory gap between the felt quality of experiences and their physical realization. In this section, I will look more closely at the precise nature of the missing explanation.

The unanswered question that is identified by both Locke and Levine has this form: 'Why this, rather than any other?'. What is missing, according to both of them, is the contrastive explanation of *why this phenomenal character rather than any other* is identical to, or produced by, a certain physiological process. That Levine is missing this contrastive explanation is clear from his formulations of the problem that I quoted above:

> there seems to be nothing about C-fiber firing which makes it naturally 'fit' the phenomenal properties of pain, *any more than it would fit some other set of phenomenal properties.* (Levine 1983, 357, emphasis added)
>
> R doesn't really explain why I have one kind of qualitative experience . . . *and not the other.* (Levine 1983, 358, emphasis added)

The same thought is present in Levine's reference to Locke, and the suggestion that corpuscular processes and ideas are stuck together in an arbitrary manner: 'The simple ideas go with their respective corpuscular configurations because God chose to so attach them. *He could have chosen differently*' (Levine 1983, 359, emphasis added). In other words, there is no explanation that would make it intelligible for us why God attached this rather than another idea to the physical process. The lack of such contrastive explanation, at least on the face of it, suggests that phenomenal characters and physiological processes are wholly different kinds of existents. They don't fit into the same order of beings, but instead they exist side-by-side.

This, according to Levine and many other commentators, is merely suggestive at this point. One way to pursue the matter is to develop the argument for the explanatory gap into *an argument for an ontological gap*. This is the anti-physicalist route.

One alternative is a certain type of physicalist view, which acknowledges that there is an explanatory gap, but resists the further development, for example, by criticizing the anti-physicalist's positive argument. Some of these physicalists felt that they still owe an account of *why* there is an explanatory gap, since the gap seems to point towards distinctness where no real distinctness exists (Papineau 2002).

In what follows, I would like to pursue a strategy that's different from both the above physicalist and anti-physicalist strategy. I want to suggest that, contrary to what Levine and Locke say, in many cases, there *is* a contrastive explanation, and the gap can be closed, or it is at least not as wide as it first appears. The explanation is provided in the framework of the causal theory of the mind. This, in itself, is not a supporting positive argument for physicalism. Just as asserting the explanatory gap does not amount to the assertion of ontological distinctness, denying the explanatory gap does not amount to denying ontological distinctness.

The 'why this, rather than that?' is a particular version of an explanatory demand, and there are other possible explanatory demands. For example, we could ask: 'why anything, rather than nothing?' That is, why do some physiological processes constitute, or give rise to conscious experiences *at all*? Another possible question: 'why this, rather than something slightly different?' Why does this physiological process give rise to a sensation of precisely this shade of phenomenal red, rather than of a somewhat darker shade?

The focus of this chapter is the first kind of question. I believe this is the question asked in Levine's paper, as the emphasis in the quotes above show. The original statement of the explanatory gap was a plea for a missing 'why this, rather than that?' explanation, and this is the explanation I aim to give. As I will clarify below, my strategy may not be suitable for satisfying the other two explanatory demands. What is achieved by my argument, if these other questions are left unexplained? In Section 9, I will return to this question and try to answer what will have been achieved by then. Here I simply ask the reader to keep in mind that I am deliberately targeting the first type of question, and not the second two.

6. The Contingency of the Realizer

Levine claims that we don't have an intelligible explanation of why the qualitative feel of pain, rather than another phenomenal property, goes together with C-fibre firing. In contrast, I would like to argue that we do have such an explanation. The explanation is based on two crucial ideas: the first is the causal theory of the mind, and the second is a thesis about the natural fit between the phenomenal character of experiences and their functional roles. This section expounds the first part, Sections 7 and 8, the second.

The first step of the causal theory, as we have seen, is to identify the causal (or functional) role of a mental phenomenon. Pain is normally caused by some damage to the body, and it normally causes avoidance behaviour. Of course, the actual causal role of pain is much more complex, and is relative to other mental states, but these can be included in the analysis.

Once we identified the functional role of an experience, we will look for the mechanism or process that plays this functional role in human beings. Here philosophers tend to behave with a certain amount of nonchalance about the empirical details, assuming that scientists will identify some or other process. Kripke himself admits that he knows virtually nothing about C-fibres, except that their stimulation is said to correlate with pain (Kripke 1972, 149). One thing we cannot take for granted is that there will be a single mechanism or process type. For example, it seems very likely that pain has a sensory and an affective component, and these can be dissociated through selective damage (Grahek 2011). Even the sensory component is likely to be realized by more than one type of process. I'm going to appeal to the same device as Levine does when he calls the 'physical story' of seeing red 'R': I am going to call the 'physical story' for pain 'C-fibre firing', and note that this story may turn out to be quite complex.

The connection between C-fibre firing and the causal role of pain involves some contingency, in two directions. First, there is the widely accepted thesis of multiple realizability, meaning that processes other than C-fibre firing could play the pain-role. The connection is not entirely arbitrary: it's not true that just any physical or physiological process could do the job. But there is almost certainly a lack of necessary connection. The other direction of contingency concerns the functions that C-fibres could play. Plausibly, they could play roles other than the pain-role—again, not just any role, but probably more than one. This is not news: most philosophers accept that the connection between the role and realizer of mental functions is, to some extent, contingent.

Levine himself is of course aware of this: at the beginning of the paper, he mentions the functionalist proposal that 'the intuition that pain could exist without C-fibers is explained in terms of the multiple realizability of mental states' (Levine 1983, 355). Since he thinks the functionalist will have other problems, he abandons this explanation. This, I believe is a mistake; what is partly (only partly, but still) responsible for the appearance of an explanatory gap for 'Pain is C-fibre firing' is an attempt to connect directly the realizer of the role with the phenomenal property.

When Levine asks what it is about C-fibre firing that fits the phenomenal properties of pain, the functional role momentarily drops out of the picture. The direct connection between C-fibre firing and pain would be difficult to establish. Levine says that the lack of intelligible connection between C-fibre firing and pain explains why in some other arguments, philosophers rely on the possibility of imagining one without the other. Continuing the presentation of Locke's view that the ideas and corpuscular processes are attached only by God's will, he says that: 'so long as the two states of affairs seem arbitrarily stuck together this way, imagination will pry them apart. Thus, it is the non-intelligibility of the connection between the feeling of pain and its physical correlate that underlies the apparent contingency of that connection' (Levine 1983, 359). But this diagnosis can be questioned. I propose to return to the familiar functionalist point: that the 'apparent contingency' is a consequence of the contingent connection between C-fibre firing and the pain-role. Even if there was some intelligible connection between the feeling of pain and the *pain-role* (as is proposed in the Section 7), the realizer would still occupy this role contingently. This is the first step in closing or narrowing the explanatory gap.

7. The Natural Fit Between Qualitative Character and the Functional Role

The first part in narrowing the explanatory gap was to identify the functional role of an experience and look for the physical process that plays this role. The second part is to point out that there is a *natural fit between the phenomenal character and the functional role of conscious experiences.*

If we follow Levine, the explanatory gap opens when we ask why C-fibre firing is identical to (or gives rise to, or is correlated with) pain, rather than with the experience of pleasure, or the experience of smelling gasoline. As we saw in Section 6, this isn't really the question we should be asking. The important question is rather this: 'Why does an experience with the phenomenal character of pain (rather than with the character of feeling pleasure, or the character of smelling gasoline) *play the pain-role*?' Levine claims that we have no intelligible explanation for this either. I think he is wrong: the answer is that the phenomenal character of pain is eminently suitable for an experience that plays this functional role.

First, pain is inherently unpleasant. This is part of the phenomenal character of pain. That pain is unpleasant makes it no surprise that it causes avoidance behaviour. Other things being equal, we prefer not having unpleasant experiences, so we try to avoid them or stop them once we have them. (If pain has a distinct sensory and affective component, then unpleasantness is part of the affective component. And that's precisely the component that seems to give rise to avoidance behaviour; see Grahek (2011).) Second, it is good to try to avoid or stop pain because pain alerts us to a damage in the body and—other things being equal—we would like to avoid or stop damage to the body. So it is a good idea that damage in the body causes this unpleasant experience which, in turn, causes avoidance. Third, physical pains normally have a felt location in the body. The felt location is part of the phenomenal character of the experience, and it directs the agent towards the possibly damaged bodily part. Fourth, by and large, more serious damage causes more intense pain, which is again part of the phenomenal character of pain.

Hence far from being arbitrary, it was in fact a very wise decision by God to choose the unpleasant experience of pain which is felt in a certain part of the body to accompany the state that is caused by damage in that part of the body, and causes avoidance behaviour.

If God is all powerful, maybe he could have associated the experience of pleasure or of smelling gasoline with the functional role of pain, but it is really hard to see how that would have worked for us. First, it's hard to see how the experience of smelling gasoline could play the role of alerting to damages in various parts of our body. Smells don't have a felt location in the way pains do. Perhaps it could be suggested that we could have different smells associated with damage in different parts of the body. That means the smell of gasoline alone is already disqualified, since it doesn't have the complexity that pain sensations do. To inform us of damage of the many different parts of the body, we would need a lot of different smells. Integration would be a problem. We are not very good in distinguishing different parts of a complex smell; if we fell and hurt our knee and hand, and we would have the sensations of the smell of onion and of burnt toast (supposing these are the simple ideas arbitrarily associated with damage in the knee and hand), it's not clear that we could discern the elements of the complex resulting sensation. They would also have to be integrated with sensations of pressure and the sensation of temperature. The whole enterprise seems rather hopeless.

As for pleasure: what we are to imagine is that pleasant feelings would cause immediate distress and avoidance behaviour. Now, it is possible to imagine that a person exhibits such reactions: suppose someone is deeply convinced that feeling any kind of pleasure is a mortal sin. For such a person, feeling pleasure could cause immediate distress. But note that this isn't a scenario where pleasure occupies the functional role of pain: rather, this explanation simply makes the story intelligible to us relative to *our* experience of pleasure, with its ordinary functional role. We noted above the familiar point that functional roles are always relative to other mental states, and this is just another instance of that point.

But what about the other case Levine mentions, that of red and green? Levine here relies on the familiar idea that red and green sensations could be inverted. Surely, here God could have chosen to attach either of the two ideas with the physical story R. In response, note that the case of red and green are very special. It is simply not true that you could invoke any other pairs of phenomenal properties and easily imagine an inversion, while retaining functional equivalence. For example, a pain-pleasure inversion, or a pain and experience-of-smelling-gasoline inversion are not plausible at all.

One reason inversion is not plausible in these cases, as mentioned above, is that different affective elements—unpleasant, pleasant, neutral—are associated with each experience, and these are plausibly responsible for different effects the experience has on our behaviour. The other reason why many inversions are not plausible is that phenomenal properties are part of a quality-space, with certain structural features (Clark 1993). Pains vary in intensity, in felt location, and in qualities such as throbbing, dull-aching, burning, or stabbing. These are all phenomenal features, and they are responsible for the felt similarities and differences among episodes of pain experiences. The structure of the quality space shows a natural fit with the typical causes of the experience, and it is reflected in our discriminatory behaviour in responding to these causes. (The concept is presented in much more detail in David Rosenthal's Chapter 6 in this volume.)

The structure of the pain quality space is quite different from the structure of the odour quality space, which in turn is different from the structure of the colour quality space and the sound quality space. So, it's hard to see how we could replace one quality by a completely different one while keeping the same structure of causes and the same range of behavioural manifestations. In fact, it was suggested that not even the red-green inversion is made possible by the phenomenal structure of colours, and that this has relevance for the explanatory gap (Hardin 1987).

8. Phenomenal Functionalism

The homomorphism between quality spaces and physical causes, and the fact that quality spaces are reflected in discriminatory behaviour, suggested to some that we could offer a reductive theory of phenomenal qualities. David Rosenthal, in Chapter 6 in this volume, presents exactly this proposal (see also Clark 1993; Rosenthal 2005). But we need not assume such a reductive view to make use of the explanatory potential of these homomorphisms. I will sketch a different background theory that goes well with the insight about the natural fit between phenomenal characters and functional roles. (Space doesn't permit me to defend this view here in detail, so I will restrict myself to its presentation.)

I propose a view which may be called Phenomenal Functionalism. It is meant to be parallel to the Phenomenal Intentionality proposal (Horgan and Tienson 2002; Loar 2003), on which the intentional features of conscious experiences are grounded in their phenomenal character. Similarly, I put forward that *(some of) the functional features of conscious experiences are grounded in their phenomenal character*. We try to avoid unpleasant experiences and seek out pleasant ones. Phenomenal qualities are located in quality spaces, and this makes them suitable to be responses to a homomorphic structure of causes which characterizes a certain portion of reality. The structure of these spaces is quite different for different kinds of qualities (for example, in different perceptual modalities), hence the functional roles of experiences—their typical causes, and the discriminatory consequences in behaviour—will be quite different. For these reasons, seamless inversions of qualities while we keep functional or physical states the same are very rare.

In addition to bodily sensations, Phenomenal Functionalism works well for occurrent moods like anxiety or elation. Anxiety is characterized by a characteristic behaviour: for example, the inability to concentrate, the tendency to get irritated, to jump on unexpected stimuli, and so on. These reactions are different from the characteristic behavioural patterns accompanying, for example, carefree contentment (Crane 1998). And the way these moods feel seems to be intrinsically connected to these roles: as in the case of pain and pleasure, it's difficult to imagine an inversion of the feeling of anxiety and contentment.

The qualitative feel of experiences is often characterized as something that's left once we extract their intentional and functional features. Tim Crane calls this the 'residue conception' of consciousness (2018). Levine seems to be suggesting something along these lines. After he acknowledges that the causal role of pain is a crucial part of our concept of pain, he says that 'there is more to our concept of pain than its causal role, there is its qualitative character, how it feels; and what is left unexplained by the discovery of C-fibre firing is *why pain should feel the way it does*!' (Levine 1983, 357, emphasis in the original).

The idea that the qualitative feel is 'more' than the causal role can be interpreted in different ways. It could mean that the qualitative feel is not reducible to the functional role, and in this sense, I am in agreement. But it could also mean that the phenomenal character is completely independent of the functional role—that once the functional role is fixed, any kind of feeling could be added. As I explained above, I find this kind of independence very implausible, but this is exactly what seems to be suggested by Levine when he talks about the explanatory gap.

The Phenomenal Intentionality programme was born out of a dissatisfaction between the sharp separation between the phenomenal and the intentional features of conscious experiences. In opposition to this separation, defenders of the programme suggest that the intentionality of mental states is inextricably linked to their phenomenal character. Phenomenal Functionalism would be a similar protest against the divide between phenomenal and functional features, and an affirmation of their deep connection.

The Phenomenal Intentionality programme is one way of opposing the sharp separation between phenomenal and intentional features. Another is representationalism, a view that also sees a strong connection between the two, but reverses their order of priority. Instead of holding that intentional features are constituted by phenomenal features, it asserts the supervenience of the phenomenal on the representational. Phenomenal Functionalism also has such a counterpart: good old functionalism. This form of functionalism would not try to eliminate phenomenal features, but would also deny the residue conception of consciousness.

Both functionalism and Phenomenal Functionalism immediately entail that there is a natural fit between the phenomenal features of an experience and its functional role. Hence they are good background theories for the explanatory task carried out in Section 6.

My sympathies lie with Phenomenal Functionalism, and someone may ask how this background theory fits into the overall project of closing the explanatory gap. Though as far as I can see, Phenomenal Functionalism, at least as characterized so far, is neutral on the issue of physicalism, it is most plausibly seen as opposed to a reductive view of phenomenal features (just like the Phenomenal Intentionality programme). One may ask whether this aspect would render its services useless in the current project. The thought is that closing the gap is a project for physicalists. A view that posits phenomenal features as basic in the explanatory theory will hardly serve this purpose.

On the picture suggested by Levine, by Locke, by the residue conception of consciousness, by the epiphenomenal qualia view, there is absolutely no intelligible connection between the functional and phenomenal features of experiences. These two aspects of reality exist side-by-side. Even if the explanatory gap is not in itself an argument for anti-physicalism, its existence is meant to be a concern for physicalists. Consequently, we may expect that anti-physicalists would cheerfully acknowledge the existence of the gap. I've been arguing in this chapter that this would be wrong, on independent grounds: whether physicalism is correct or not, the sharp separation of the phenomenal and the functional is simply implausible. Our experiences fit very well into the physical world and our place in it. Instead of arbitrarily assigning simple ideas to physical processes, God, in his wisdom, assigned phenomenal characters to experiences that are eminently suitable for playing certain functional roles.

9. Various Gaps

Conscious experiences have a phenomenal character, and this character has a natural fit with the functional role of the experience. Hence the explanatory demand 'why this, rather than that?', posed by Levine, can be satisfied. C-fibre firing correlates with the experience of pain, rather than, say, pleasure, because in human beings, C-fibre firing occupies the functional role of pain, and the phenomenal character of pain naturally fits this functional role (much better than it would fit the functional role of pleasure).

Thus, we can answer the 'why this, rather than that?' question. Where does this leave us with the other possible explanatory questions we identified in Section 4? Let us first consider the question 'why this, rather than something slightly different?' I am not sure what to say about this case. If we identify the functional role of experiences on the basis of a common-sense analysis, then it seems that there is no answer to this question. Yes, the unpleasantness of pain is suitable for causing avoidance behaviour, but a slightly more or less intense pain would also seem suitable to cause the same. Yes, these causes and this discriminatory behaviour perfectly fit the inner similarity relations among the sensations of shades of red, but if all shades felt a touch darker, the fit would seem to be the same. This could suggest that the strategy pursued here cannot answer this explanatory question.

However, what makes me hesitate is the prospect of a scientific inquiry, for example, into quality spaces and their corresponding causes, that would reveal subtle differences undetected by common-sense analysis. Something like this happened in the case of red-green inversion: common sense seems to license the possibility of such an inversion, but it has been argued that a more subtle investigation of the colour quality space shows that a perfect functional equivalent is not possible (e.g., Hardin 1987). Maybe more subtle investigations would also exclude the possibility of shifts.

Can we answer the 'why anything, rather than nothing?' question—in other words, can we explain why some physiological processes give rise to conscious experiences at all? According to the line of thought presented in this chapter, once we have the functional role of states or events on one side, and the phenomenal characters on the other, we can pair them according to the natural fit between character and role. But the same strategy does not explain why there were phenomenal characters in the first place. Some of our bodily functions seem to work perfectly well without the involvement of consciousness. For example, arguably, processes that occur in the dorsal visual stream, which takes visual stimuli as input and issues motor commands as an output, do not give rise to consciousness (Goodale and Milner 1992). In contrast, processes in the ventral stream do give rise to consciousness. The strategy pursued here does not explain why.

In fact, this challenge can also be divided to further, more specific tasks. One is to give an account of the general function of consciousness (for example, having to do with combining information from several sources), and try to answer questions like the one above. When certain interactions with the world fell in the category of requiring consciousness, we could detect this from their functional role. This theory would assign a uniform account for all conscious processes.

There is a different task that Mary, the scientist in the black and white room faces (Jackson 1982; Howard Robinson 1982 has a similar example with a deaf scientist). Let us first consider a modified Mary story, where Mary lives in a normal world and is accorded the full range of experiences, but her studies are curiously detached from her experiences. She studies in detail the physics and physiology of sensations, perceptions, and emotions, but she is not told which of her experiences are connected to which physiological process. For example, she studies 'C-fibres' (or whatever complex physiological process which underlies pain), but is never told that this process in fact underlies pain. According to the argument pursued in this chapter, after a while, she would be able to make the connection herself.

Original Mary is in a less favourable position. Her task is not to pair experiences with physical processes, but rather to somehow deduce the nature of the experience from the physical process, without any further help. The consensus is that she won't be able to do that. She may be able to form some idea: she will know it's a colour experience, it's not an emotion or a bodily sensation, and since she has had those, she will have some reasonable expectations of the phenomenal character of experience. But she won't know it exactly.

One immediate consequence is that if there is an epistemic gap between the physical and the phenomenal, it will be the kind of gap that we find in original Mary's understanding. Indeed, Levine himself, in his book *Purple Haze* (Levine 2001), shifts focus to the absent qualia argument and the knowledge argument from the kind of explanatory question that he considers in the original paper. Note, however, that Mary's predicament is not very naturally characterized as lacking some *explanation*. Explanations are usually answers to 'why' questions; they have an explicandum: a fact, or an event. Mary's task is to deduce the character of the experience from its physical description. It doesn't look like she is trying to answer a why-question, or that there is a fact or event that she is trying to explain. In contrast, in Levine's 1983 paper, there was a clear why-question waiting for an answer: 'why this, rather than that?' There is also the other general explanatory question—why are some processes conscious?—but that is not the question Mary is trying to answer either.

So, while Mary has a gap in her understanding or knowledge, I would be reluctant to call it an 'explanatory gap'. In any case, no matter what we call it, the question is whether this gap threatens our understanding of the mind. The answer will partly depend on our assessment of what is achieved by explanations about physical phenomena.

For example, do we have an explanation of why the fundamental constituents of the universe are what they are? Couldn't we have a world with a slightly different assortment of elementary particles? It is arguable that certain natural constants that determine the strength of interactions, have, according to our best theories, taken spontaneous values during the evolution of the universe, and could have had other values within a certain range. And many of these facts affect the world of the elementary parts. Perhaps it's just a fact about our world that it contains certain basic ingredients. Similarly, we could argue that it is a fact about our world that it contains certain conscious experiences.

If this is right, then answering Levine's question, 'why this, rather than another?' actually goes a long way to close or narrow the explanatory gap. We have a world where certain beings have conscious experiences with phenomenal characters. These phenomenal characters don't belong to a realm that is entirely disconnected from the physical realm: in fact, there is a natural fit between the two kinds of features. There isn't a huge gap between the physical and phenomenal aspects of the world.

References

Armstrong, D. M. (1968) *A Materialist Theory of Mind*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1993) 'Preface', in *A Materialist Theory of the Mind*, 2nd edition, London: Routledge, xiii–xxiii.

Armstrong, D. M. (1999) *The Mind-Body Problem: An Opinionated Introduction*, Boulder, CO: Westview Press.

Berkeley, George (1949) *A Treatise Concerning the Principles of Human Knowledge*, in *The Works of George Berkeley*, vol. 2, edited by A. A. Luce and T. E. Jessop, London: Thomas Nelson and Sons, 19–113.

Clark, Austen (1993) *Sensory Qualities*, Oxford: Oxford University Press.

Crane, Tim (1998) 'Intentionality as the Mark of the Mental', in *Royal Institute of Philosophy Supplement*, vol. 43, Cambridge: Cambridge University Press, 229–51.

Crane, Tim (2018) 'A Short History of the Philosophy of Consciousness in the Twentieth Century', in *Philosophy of Mind in the Twentieth and Twenty-First Centuries: The History of the Philosophy of Mind*, vol. 6, edited by Amy Kind, London: Routledge, 78–103.

Goodale, Melvyn A. and A. David Milner (1992) 'Separate visual pathways for perception and action', *Trends in Neurosciences*, 15: 20–25.

Grahek, Nikola (2011) *Feeling Pain and Being in Pain*, Cambridge, MA: MIT Press.

Hardin, Clyde L. (1987) 'Qualia and Materialism: Closing the Explanatory Gap', *Philosophy and Phenomenological Research*, 48: 281–98.

Horgan, Terence and John Tienson (2002) 'The Intentionality of Phenomenology and the Phenomenology of Intentionality', in *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers, Oxford: Oxford University Press, 520–33.

Jackson, Frank (1982) 'Epiphenomenal Qualia', *Philosophical Quarterly*, 32: 127–36.

Kripke, Saul A. (1972) 'Naming and Necessity', in *Semantics for Natural Language*, edited by Donald Davidson and Gilbert Harman, Dordrecht: Reidel. Reprinted in book form, Oxford: Blackwell, 1980.

Levine, Joseph (1983) 'Materialism and Qualia: The Explanatory Gap', *Pacific Philosophical Quarterly*, 64: 354–61.

Levine, Joseph (2001) *Purple Haze: The Puzzle of Consciousness*, New York: Oxford University Press.

Loar, Brian (2003) 'Phenomenal Intentionality as the Basis of Mental Content', in *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, edited by Martin Hahn and B. Ramberg, Cambridge, MA: MIT Press, 229–58.

Locke, John (1975) *An Essay concerning Human Understanding*, edited by Peter H. Nidditch, Oxford: Clarendon Press.

Papineau, David (2002) *Thinking about Consciousness*, Oxford: Clarendon Press.

Robinson, Howard (1982) *Matter and Sense: A Critique of Contemporary Materialism*, Cambridge: Cambridge University Press.

Rosenthal, David (2005) *Consciousness and Mind*, Oxford: Oxford University Press.

9

Introspection and Distinctness

Armstrong and Shoemaker on Introspection

Ryan Cox

Claims about the distinctness or non-distinctness of introspective beliefs from the mental states they are about have played a central role in the philosophy of introspection in the past fifty years or so. In *A Materialist Theory of the Mind* and work leading up to it, David Armstrong famously argued against infallibilist theories of introspection, and in defence of his own self-scanning theory of introspection, on the grounds that introspective beliefs are distinct from the mental states they are about. Sydney Shoemaker, one of Armstrong's most ardent critics, famously argued against Armstrong's self-scanning theory of introspection, and in favour of his own constitutive theory of introspection, on the grounds that introspective beliefs are not distinct from the mental states they are about. Yet the relevant sense or senses of distinctness involved here, and the role such claims about distinctness play in such arguments, are notoriously hard to pin down. This chapter explores some of the issues concerning distinctness and non-distinctness in the philosophy of introspection and in the dispute between Armstrong and Shoemaker and offers a reassessment of some of the central arguments offered in that dispute.

1. Introduction

Among the distinctive contributions of *A Materialist Theory of the Mind* is a particularly clear statement of a theory of introspection, a theory I will call, for reasons which will become apparent, the self-scanning theory of introspection. In offering this theory, Armstrong took himself to be continuing in the tradition of Locke, Kant, and Brentano, and to be defending what he took to be part of the traditional theory of mind.[1] According to the traditional theory, the model for

---

[1] Of Locke and Kant, Armstrong writes: 'Locke spoke of the faculty of reflection, Kant spoke of inner sense, modern philosophy often speaks of introspection. I believe this traditional view to be essentially correct' (Armstrong 1963, 417). See also Armstrong (1980b, 61; 1984a, 109–10). Armstrong claims that Brentano is a proponent of the traditional theory (Armstrong 1984a, 109) and at one point calls the traditional theory the 'Locke-Kant-Brentano theory' (Armstrong 1984a, 113).

Ryan Cox, *Introspection and Distinctness: Armstrong and Shoemaker on Introspection* In: *Armstrong's Materialist Theory of Mind.* Edited by: Peter R. Anstey and David Braddon-Mitchell, Oxford University Press. © Ryan Cox 2022. DOI:10.1093/oso/9780192843722.003.0010

introspection is ordinary sense-perception and introspection is seen as a kind of inner sense-perception.[2] In a widely quoted passage, Armstrong writes:

> I believe that Kant suggested the correct way of thinking about introspection when he spoke of our awareness of our own mental states as the operation of 'inner sense'. He took sense-perception as the model for introspection. By sense-perception we become aware of the current physical happenings in our environment and our body. By inner sense we become aware of current happenings in our own mind. (*MTM*, 95)

Armstrong's distinctive contribution to the traditional theory was to combine it with the central-state theory of mind—according to which, mental states are states of the central nervous system, or more particularly, of the brain—and so to arrive at the view that introspection is a 'self-scanning process in the brain' (Armstrong 1963, 419).[3]

Armstrong's nod to Locke, Kant, and Brentano, and to the traditional theory, is problematic on several fronts and, unfortunately, it obscures the distinctiveness of his contribution. Take the nod to Locke, for instance. As Sydney Shoemaker points out, Locke didn't take introspection to be like sense-perception as Locke himself took the latter to be, rather, he took it to be like sense-perception as we naively take it to be, that is, as providing us with immediate, direct, access to its objects.[4] It is important, in understanding Armstrong's claim that introspection is like sense-perception, to note that Armstrong thought that introspection was like sense-perception *as he took it to be*, not as we naively take it to be—and so not as Locke took it to be—and not as many contemporary philosophers take it to be.[5] Many philosophers think that the perceptual access we have to external objects is *not* as Armstrong took it to be. Many philosophers take sense-perception to involve perceptual states distinct from perceptual beliefs, whereas Armstrong took perception merely to involve the getting of beliefs without any intervening perceptual states.[6] It is a distinctive feature of the self-scanning theory that it has no place for introspective states modelled on perceptual states distinct from introspective beliefs. So, the self-scanning theory is better thought of as a completely new theory of introspection, one without historical precedent.

---

[2] According to Armstrong, the traditional theory 'asserts that for the case of a person's own mind, though not for any other mind, a perception-like process goes on which yields the person information about the current state of his mind' (Armstrong 1984a, 108–9). Armstrong took bodily perception, or proprioception, to be the kind of sense perception which provided the most fitting model for introspection: 'Reflection, inner sense, introspection, introspective awareness, consciousness (in one sense of the word), whatever we call it, can seem a very mysterious phenomenon. But I believe that there exists a demystifying, naturalizing, *model* for it in the field of perception. I refer to *bodily perception*, or, as some psychologists call it, *proprioception*' (Armstrong 1984a, 110). See also (*MTM*, 96).

[3] See also (*MTM*, 102–3) and Armstrong (1984c, 22).

[4] Shoemaker (1996b, 203).

[5] For Armstrong's views on perception see (*MTM*, Chapter 10).

[6] Moreover, Kant, Locke, and Brentano all held theories of introspection as inner-sense, where inner-sense involved sensations or sensory qualities. The self-scanning theory had no place for

For Armstrong, the self-scanning theory simply fell out of the combination of the traditional theory of introspection, according to which introspection is a perception-like process, and the central-state theory, according to which mental states are states of the central nervous system. And it is because Armstrong took the self-scanning theory to simply fall out of the combination of the traditional theory of introspection and the central-state theory that he was concerned, throughout his work in the philosophy of mind, with the threat posed by the purported infallibility of introspection.[7] Infallibility, the logical impossibility of having false beliefs about our own mental states, seemed incompatible with the combination of the central-state theory and the traditional theory of introspection. So, according to Armstrong: 'It is essential . . . for the defender of a Central-state Materialism to show that there can be no logically indubitable knowledge of, or logically privileged access to, or self-intimation by, our current mental states' (*MTM*, 103). It was in response to this concern that Armstrong developed what I will call the *argument from distinctness*, or the distinct existences argument, as it is sometimes called.[8] A central premise of this argument is that one's introspectively believing that one is in some mental state is *distinct* from one's being in that mental state. The other central premise—Hume's dictum, as it is sometimes called—is that there can be no necessary connections between distinct existences. An examination of this argument will be a central theme of the discussion to come. The consensus seems to be that Armstrong's argument is ultimately unsuccessful in establishing its conclusion since there *can* be necessary connections between distinct existences—that is, Hume's dictum is false. Nonetheless, the conclusion of the argument is widely taken to be true: we are not infallible about our own mental states. My interest in what follows will not be with the conclusion

---

sensations or sensory qualities. The question of sensations is complicated by Armstrong's discussion in *Consciousness and Causality*, however. Armstrong says that his model for inner-sense is bodily perception or proprioception (Armstrong 1984a, 110). He writes 'over and above the five senses, there is awareness of bodily temperature, awareness of position and movement of limbs, and awareness of the relation of the body to the earth. We bring these forms of awareness together as bodily perception, or proprioception' (Armstrong 1984a, 110). And he says that, 'it is in many cases difficult to pick out the sensation involved in bodily perception' (Armstrong 1984a, 111). But he goes on: 'However, I think that in fact sensations are always involved' (Armstrong 1984a, 111). These remarks open Armstrong to the objection that introspection is not like perception, since perception always involves sensations or sensory qualities and introspection involves no such thing. Lycan replies on behalf of the inner-sense theorist to such an objection: 'The inner-sense theorist does not contend (at least neither Armstrong nor I contend) that internal monitoring is like external perception in every single respect. And in particular, we should not expect internal monitoring to share the property of involving some presented sensory quality at its own level of operation' (Lycan 1987, 28).

[7] Armstrong uses the term 'incorrigible' for 'infallible' in his earlier work. See Armstrong (1963) and (*MTM*, 100). As Armstrong notes, in rejecting the infallibility of introspection, he is rejecting a view he once held in Armstrong (1961, Chapter 4) and Armstrong (1962, Chapter 9). See Armstrong (1963, 418) and (*MTM*, 103).

[8] See Jackson (1973, 57).

about fallibility. Rather, my interest will be with Armstrong's defence of the premise that one's introspectively believing that one is in some mental state is distinct from one's being in that mental state. If Armstrong has a good argument *for* this premise, then he has a good argument *against* a central premise in a particular argument *against* the self-scanning theory of introspection.

The self-scanning theory, like inner-sense theories of introspection more generally, has few defenders today.[9] It is either seen as essentially tied to a problematic form of reliabilist epistemology, or as being empirically implausible.[10] Neither the epistemological nor empirical objections strike at the heart of the self-scanning theory, however.[11] For an objection that attempts to strike at the heart of the theory, we need to turn to Sydney Shoemaker's arguments against the self-scanning theory, and inner-sense theories in general.[12] According to Shoemaker, it is an undeniable commitment of the self-scanning theory that my introspectively believing that I am in some mental state is distinct from my being in that mental state. Shoemaker argues against the self-scanning theory exactly on the grounds that my introspectively believing that I am in some mental state is *not* distinct from my being in that mental state. This is Shoemaker's central argument against the self-scanning theory and the inner-sense theory, and there is a widely held view in the literature on self-knowledge that Shoemaker has decisively refuted the self-scanning theory. For instance, Fred Dretske writes: 'Sydney Shoemaker... gives what I regard as a definitive critique of perceptual models of self-knowledge' (2003, 11).[13] But for such a decisive and definitive objection, Shoemaker's objection is notoriously difficult to pin down and the tide seems to have turned against Shoemaker's arguments, at least among some theorists.[14] So, a second theme of the discussion to follow will be what I call Shoemaker's *argument from non-distinctness*, what is sometimes called the argument from self-blindness.

The following discussion will be centred around the reconstruction of Armstrong's argument from distinctness and around Shoemaker's argument from non-distinctness. Such reconstructions require careful attention to the sense or senses of 'distinctness' involved in the respective arguments. The overall narrative of the discussion will be as follows. Armstrong's argument from distinctness against infallibility is ultimately unsuccessful, as he admits, because he is unable to establish the premise that my introspectively believing that I am in some mental state is distinct from my being in that state.[15] With the distinctness premise hanging, we take up Shoemaker's argument that my introspectively believing that I am in some mental state is *not* distinct from my being in that mental state and so the self-scanning theory is false. I argue that, understood as a lowly argument to the best explanation, Shoemaker's argument for this conclusion may well go through. So, the self-scanning theory, as Armstrong understands it, is in trouble. I close with some remarks about the depth—or perhaps indeed the shallowness—of the disagreement between Armstrong and Shoemaker on the matter of distinctness.

---

[9] Although see Lycan (1987, 13, 72–4; 1996, 16, 31); and Nichols and Stich (2003, 161–2).

[10] For the epistemological objections, see Zimmerman (2008); Gertler (2011); and Dretske (2012). For the empirical objections, see Lyons (1986, 55); and Carruthers (2011, 194–5).

[11] The epistemological objections often rely on controversial epistemological theses and the empirical objections often involve highly contestable findings and inferences. Moreover, similar epistemological objections can be raised against most alternative theories of introspection, including the theories of critics of the self-scanning theory. So epistemological objections are often moot.

[12] Shoemaker (1996a; 1996c; 2012).

[13] Adding a further epistemological objection in another context, Dretske writes 'I take this to be merely an additional (probably unnecessary) nail in the coffin Shoemaker (1996) constructed for perceptual models of introspection' (Dretske 2012, 59, n. 12). Thanks to Donald Nordblom for reminding me of this passage.

[14] See Kind (2003); and Gertler (2011).

2. The Argument from Infallibility

To put the argument from distinctness into context, we have to examine another argument. The argument is an argument in response to which Armstrong offers the argument from distinctness, an argument I will call the *argument from infallibility*. Infallibility, in this context, is the claim that it is logically impossible to introspectively believe that you are in a particular mental state without being in that mental state. It is this claim which Armstrong took to be a threat to the central-state theory, since, on the face of it, infallibility seems to be incompatible with the central-state theory. This section examines the argument from infallibility and introduces an important theme for the discussion to follow: throughout Armstrong's discussion of introspection and the central-state theory, Armstrong simply takes the traditional theory of introspection for granted and is concerned primarily with the opposition between the central-state theory and the forms of dualism he is seeking to displace.

The argument from infallibility first appears in 'Is Introspective Knowledge Incorrigible?' (Armstrong 1963, henceforth IKI).[16] Armstrong writes:

> I wish to defend the thesis, recently advanced by J. J. C. Smart, and others, that mental states are, as a matter of contingent fact, states of the brain. Now if I accept the existence of introspection, as I also do, then I must conceive of both introspection and the objects of introspection as states of the brain. Introspection must be a self-scanning process in the brain. That it is logically possible that such a self-scanning process will yield wrong results is at once clear... So if

---

[15] Armstrong ultimately leaves the distinctness question open to empirical hostage in response to a suggestion of David Sanford, discussed in Section 3.4.

[16] Armstrong cites Kurt Baier as the source of this line of argument (*MTM*, 103), although it is quite difficult to recover an argument along these lines from Baier (1962).

> introspection is incorrigible... it seems that a materialist doctrine of mental states must be false. (Armstrong 1963, 418–19)

The argument appears in basically the same form in *A Materialist Theory of the Mind* (henceforth *MTM*). There Armstrong writes:

> If mental processes are states of the person apt for the bringing about of certain sorts of behaviour, and if these states are in fact physical states of the brain, then introspection itself, which is a mental process, will have to be a physical process in the brain. It will have to be a self-scanning process in the brain. Now it is at once clear that it is always logically possible, at the very least, for such a self-scanning mechanism to yield the wrong result. For any mechanism can fail to operate properly. So if introspective knowledge is incorrigible, as is alleged, then Central-state Materialism is false. (*MTM*, 102–3)[17]

A cursory reading of these passages, especially the latter, suggests that Armstrong has something like the following argument against the central-state theory in mind.

The Argument from Infallibility*

P1. If the central-state theory is true, then the self-scanning theory of introspection is true.

P2. If the self-scanning theory of introspection is true, then we are not infallible about our own mental states.

P3. We are infallible about our own mental states.

C1. The central-state theory is false.

Something like P1 is strongly suggested by the first two sentences of the quoted passage from *MTM*. P2 of the reconstruction is strongly suggested by the two sentences after this one. And P3 and C1 are strongly suggested by the final sentence.

This reconstruction of Armstrong's argument would make sense of other aspects of Armstrong's discussion. After all, the section in which the argument occurs is called 'The contradiction with Central-state Materialism', and the contradiction here is between infallibility and the central-state theory. Armstrong also writes: 'It must now be shown that these doctrines of incorrigible awareness of, logically privileged access to, and self-intimating nature of, our own current mental states entail the falsity of Central-state Materialism' (*MTM*, 102). But if this reconstruction is the argument that Armstrong had in mind, then it is a wonder why he took it so seriously. There is no reason to think that if the central-state theory is true, then the self-scanning theory is true. That is, there is no reason to think that P1 is true. Indeed, other philosophers of a materialist persuasion have been sanguine about the possibility of combining infallibility and something like the central-state theory.[18]

---

[17] See also Armstrong (1984c, 22).

A better reconstruction of Armstrong's argument is the following:

The Argument from Infallibility

P1. If the traditional theory of introspection is true, and central-state materialism is true, then the self-scanning theory of introspection is true.

P2. If the self-scanning theory of introspection is true, then we are not infallible about our own mental states.

P3. We are infallible about our own mental states.

C1. If the traditional theory of introspection is true, then central-state materialism is false.

This reconstruction more closely parallels Armstrong presentation in the quoted passage from IKI. Something like P1 is strongly suggested by the first three sentences of the passage quoted above. Notice there that Armstrong takes his commitment to introspection—that is, to the traditional theory of introspection—as an additional commitment. It is the combination of the traditional theory of introspection and the central-state theory which entails the self-scanning theory. Notice, however, that in order to make the argument valid, the conclusion must take a conditional form. The contradiction, then, is between the central-state theory, the traditional theory of introspection, and infallibility. Armstrong's claim should then be, 'It is essential . . . for the defender of a Central-state Materialism [who is also committed to the traditional theory of introspection] to show that there can be no logically indubitable knowledge of, or logically privileged access to, or self-intimation by, our current mental states' (*MTM*, 103). The structure of Armstrong's argument here reveals the sense in which the traditional theory of introspection is being presupposed between Armstrong and his opponents.

Let us grant the truth of P1, since Armstrong all but makes it definitional of the self-scanning theory that it is entailed by the central-state theory and the traditional theory of introspection. In support of P2, Armstrong writes: 'Now it is at once clear that it is always logically possible, at the very least, for such a self-scanning mechanism to yield the wrong result. For any mechanism can fail to operate properly' (*MTM*, 102–3). Let's also grant the truth of P2 then: if self-scanning is a perception-like process, and it is a process in the central nervous system, then, arguably, the self-scanning mechanism can yield the wrong result and we are not infallible about our own mental states.[19] If we grant both assumptions, and we are committed to the traditional theory of introspection, then the central-state theory looks to be in trouble. It is no surprise then, that it is P3 of the argument from infallibility against the central-state theory that Armstrong takes issue with. Such a thesis can be found in traditional foundationalist epistemology and empiricist theories of mind. But Armstrong admits that he can see no serious arguments *for* it.[20] Nonetheless, he feels the need to offer an argument against it. Armstrong's argument against the infallibility premise is the argument from distinctness against infallibility, which we will turn to in Section 3.

---

[18] See Lewis (1972, 258).

3. The Argument from Distinctness

Armstrong gives several arguments against infallibility, but the one we will be interested in here is the argument from distinctness.[21] The argument occurs in both IKI and *MTM*. Here is Armstrong's brief statement of the argument in IKI:

> The acquiring of introspective knowledge must consist of the making of (sincere) reports of current mental occurrences, or else a nonverbal apprehension of these occurrences. In both cases the apprehension of the occurrence will have to be *distinct* from the occurrence that is apprehended. But if this is granted, then we can apply Hume's argument about 'distinct existences'. Wherever we have two distinct things, Hume points out, there we can always conceive of the one existing in the absence of the other. It follows that it is logically possible to have a sincere report of a current inner experience, or a nonverbal apprehension of that experience, without the experience existing. ('Apprehension', of course, is a success-word, and strictly it would be out of place here. But it would be possible for there to be something just like the apprehension, except that the object of the apprehension did not exist.) But this state of affairs would be the state in which we would be mistaken about our current inner mental state.
>
> (Armstrong 1963, 422–3)[22]

In short, Armstrong wants to argue that it is logically possible for me to introspectively believe that I am in a particular mental state without being in that mental state on the grounds that my introspectively believing that I am in a particular mental state and my being in that mental state are distinct existences.

---

[19] The premise has been challenged by Pappas (1976). For Armstrong's reply, see Armstrong (1976).

[20] See Armstrong (1984c, 124).

[21] See (*MTM*, 104–6) for the other arguments. See Jackson (1973) for a critical discussion of these arguments and others. Armstrong credits J. J. C. Smart with independently arriving at the argument; see Smart (1962).

[22] See also (*MTM*, 106).

Let *M* refer, here and throughout, to an arbitrary mental state like the state of believing that it is raining. The argument from distinctness can be represented as follows:

The Argument from Distinctness

P1. If my introspectively believing that I am in *M* is distinct from my being in *M*, then it is logically possible for me to introspectively believe that I am in *M* without being in *M*.

P2. My introspectively believing that I am in *M* is distinct from my being in *M*.

C1. It is logically possible for me to introspectively believe that I am in *M* without being in *M*.

Since we are going to have to refer back to these premises and various versions of them, let us call P1 of the argument from distinctness the no necessary connection premise, P2 the distinctness premise, and C1 the fallibility conclusion. C1 is just the denial of infallibility, so the argument from distinctness offers a response to the argument from infallibility.

A few remarks are in order about this reconstruction. First, here, as elsewhere, I have retained Armstrong's 'logically possible'. If logical and metaphysical possibility are different, then the argument might go differently if cast in terms of metaphysical possibility. As far as I can tell, it doesn't make a difference for our purposes which notion of possibility we use in the argument. Second, and more importantly, I have reconstructed the argument in terms of relations between states of affairs, namely, the state of affairs of my believing that I am in *M* and the state of affairs of my being in *M* rather than in terms of relations between mental states themselves, that is, between the state of believing that I am in *M* and *M* itself. This is natural, as Armstrong often speaks of the relata which are meant to be distinct from each other as if they were particulars, and mental states are not particulars in the relevant sense.[23] Finally, by 'introspectively believing that I am in *M*', I mean to be making a claim about how I came to believe that I am in *M*: I came to believe it by introspection, whatever that is. No assumptions are being made about the nature of introspection here. A contrast is simply being drawn with uncontroversially non-introspective ways of coming to believe things, like coming to believe things about one's own mental states on the basis of testimony.

---

[23] See Armstrong (1973, Chapter 12) for some insight into Armstrong's views on the state of affair of being in a particular mental state. I have avoided talk of 'token mental states' altogether, although it is tempting to cast the argument in terms of token mental states. I have done so because I am sceptical about talk of token mental states if it is taken to be distinct from talk about states of affairs of being in mental states.

The cogency of the argument from distinctness is going to come down to how 'distinct' is understood throughout. And it is not immediately clear how it should be understood. A common complaint here is that Armstrong doesn't say much about what he means by 'distinct'. As Frank Jackson writes, 'Armstrong never really makes clear the sense he is giving to "distinct"' (1973, 57). And as David Sanford writes: 'The phrase "distinct existences" occurs throughout Armstrong's writings but receives little explanation' (1984, 76). Both assessments seem fair.[24] Armstrong gives us little clue as to what he means by 'distinct'. He gives us one clue when he uses the modifier 'logically' in speaking of things being 'logically distinct' (*MTM*, 107). It might be thought that this settles the matter in favour of thinking that Armstrong has numerical distinctness in mind since logical identity is numerical identity. But I doubt that this settles the matter. There are other understandings of distinctness that are just as deserving of the modifier 'logical', and we will discuss some of these below.

So, we face two questions about the argument from distinctness. The first is the purely exegetical question of what Armstrong means by 'distinct', and the second is the philosophical question of whether there is a sense of 'distinct' on which the argument from distinctness is cogent. The exegetical question is complicated by the possibility that Armstrong might not have had the same sense of distinctness in mind at various places in his writing. One thing is for sure, however, and that is that Armstrong was taking his cue here from Hume. As he writes: 'Wherever we have two distinct things, Hume points out, there we can always conceive of the one existing in the absence of the other' (Armstrong 1963, 422). There is, therefore, a further exegetical question of what Hume meant by 'distinct' and 'existence', but we won't take up that question here. Since Armstrong offers no arguments for the no necessary connections premise, all we can do is ask what he means by 'distinct existence' and ask whether it is true that there are no necessary connections between distinct existences. However, since Armstrong does offer arguments for the distinctness premise, we can attempt to reconstruct and evaluate these.

Our strategy, then, in what follows, will be to consider different versions of the argument from distinctness, distinguished by the sense of 'distinct' they appeal to, and to ask whether they are cogent, and whether they represent the argument Armstrong had in mind.

3.1 The Argument from Numerical Distinctness

An obvious place to start our discussion of different notions of distinctness is with the notion of numerical distinctness. Numerical distinctness is probably the best understood notion of distinctness in the vicinity, and also the most general, for everything is numerically identical to something or it is not. In starting with numerical distinctness, I should immediately note that it is obvious that things can be numerically non-identical without being distinct. (They cannot be numerically non-identical without being numerically distinct, since to be numerically distinct is just to be numerically non-identical.) So, we should note at the outset that there is going to be at least one other sense of 'distinctness' we are going to have to explore. We define numerical distinctness in terms of numerical non-identity and derive the following as a special case:

---

[24] Although, Armstrong does say more about what he means by distinctness in later works. See Armstrong (1989, 115–18) and Armstrong (1997) for further discussion.

> **Numerical Distinctness**: my introspectively believing that I am in *M* is numerically distinct from my being in *M* if and only if my believing that I am in *M* is not numerically identical to my being in *M*.

With this understanding of distinctness in hand, we have the following version of the argument from distinctness:

> **The Argument from Numerical Distinctness**
>
> **P1.** If my introspectively believing that I am in *M* is numerically distinct from my being in *M*, then it is logically possible for me to introspectively believe that I am in *M* without being in *M*.
>
> **P2.** My introspectively believing that am in *M* is numerically distinct from my being in *M*.
>
> **C1.** It is logically possible for me to introspectively believe that I am in *M* without being in *M*.

Is the argument cogent? We can begin by examining P2, the numerical distinctness premise. On the face of it, it seems plausible to assume that my introspectively believing that I am in *M* is numerically distinct from my being in *M*, even though our intuitions about the numerical distinctness of states of affairs are a bit obscure. One way to *argue* for this claim is to claim that it is sufficient for the numerical distinctness of states of affairs that they involve distinct states or properties. Thus, we can move from the very plausible claim that the state of believing that I am in *M* and *M* itself are numerically distinct mental states to the conclusion that the state of affairs of my believing that I am in *M* and my being in *M* are distinct. If *M* is the state of believing, say, that it is raining, then it is very plausible to claim that the states are distinct, since contentful states are individuated by their contents, and these states clearly have numerically distinct contents.

The problem with this version of the argument from distinctness emerges when we examine P1, the no necessary connections premise. There seems to be an emerging consensus that there *can* be necessary connections between numerically distinct existences. This is clearest in the case of properties: the property of being crimson is necessarily connected with the property of being red. Of course, to say that they are necessarily connected is to say that something cannot have one property without having the other. But recall Hume's own way of making the point: if two things are distinct, then we can always conceive of one existing without the other. It is a somewhat obscure matter as to what it would be for the property of being crimson to exist without the property of being red existing. So, to be fair to Armstrong and Hume, we should consider examples involving particulars, and we should take Hume's dictum to be restricted to particulars. However, even if we do this, it seems that there are examples of numerically distinct states of affairs involving things being in particular states, where there are necessary connections between something being in one state and being in the other. The chair's being red and the chair's being coloured are distinct states of affairs. They are distinct since the latter is constituted by the former: that is, the chair's being coloured is constituted by the chair's being red. And yet the states of affairs are necessarily connected: it is logically necessary that if the chair is red, then it is coloured. This seems to be as clear a counter-example as any to the no necessary connection premises understood in terms of numerical distinctness. Once this example has been noted, examples can be multiplied. Indeed, there are arguably even examples involving mental states. Consider the relation between my knowing that it is raining and my believing that it is raining.[25] These are numerically distinct states of affairs and there is a necessary connection between my knowing that it is raining and my believing that it is raining. Or consider the relation between my intending to get wet and my wanting to get wet. Arguably these are numerically distinct states of affairs and yet there is a necessary connection between my intending to get wet and my wanting to get wet.

So, we have seen that the distinctness premise of the argument from distinctness is plausibly true, when understood in terms of numerical distinctness, and we have seen that the no necessary connections premise of the argument is almost certainly false, when understood in terms of numerical distinctness.

It follows that if the argument from distinctness is cogent on some understanding of 'distinct', then it must be cogent on a different understanding of 'distinct'. If Armstrong had numerical distinctness in mind when he presented the argument, as some have claimed, then the argument he presented is not cogent.[26] But charity requires us to assume that it cannot be this sense of distinctness which Armstrong had in mind. As Daniel Stoljar writes:

---

[25] See Williamson (2000) for the claim that knowledge is a mental state. The truth of this claim matters little for the counter-example.

[26] Speaking of the discussion in *MTM*, Jackson writes: 'Here it seems that Armstrong is using "distinct" to mean numerically distinct, because his argument is designed to show that subject and object cannot be one and the same thing' (1973, 59).

> It is reasonable to suppose that philosophers who invoke Hume's dictum are getting at a particular notion of distinctness. They are not making an obvious mistake. Nor are they talking about nothing. On the other hand, the notion of distinctness that is at issue in Hume's dictum just cannot be numerical distinctness, for if it were the dictum is obviously false. (Stoljar 2008, 269)

We will now consider whether the argument is cogent on a different understanding of 'distinct'. Then we will turn to Armstrong's arguments for the distinctness premise and ask whether they support the distinctness premise on some understanding of 'distinct'.

3.2 The Argument from Constitutive Distinctness

I began Section 3.1 by noting that things can be numerically non-identical without being distinct. So, there is at least one sense of 'distinctness' distinct from numerical distinctness. Indeed, there seem to be several senses of 'distinctness' distinct from numerical distinctness. I will consider what I take to be the most important alternative notion of distinctness in the vicinity, namely, constitutive distinctness.[27]

We have already considered claims of constitution between states of affairs. The chair's being coloured, I claimed, is constituted by its being red. Here we have numerically distinct states of affairs related by the relation of constitution. Similarly, John's being a bachelor, is party constituted by his being unmarried. Here again we have numerically distinct states of affairs related by the relation of constitution. Constitutive distinctness gives us another natural notion of distinctness where states of affairs can be numerically non-identical without being distinct. Constitutive distinctness is defined in terms of relations of constitution, and we derive the following as a special case:

> **Constitutive Distinctness**: my introspectively believing that I am in *M* is constitutively distinct from my being in *M* only if (i) it is not the case that my believing that I am in *M* is partly or wholly constituted by my being in *M*, and (ii) it is not the case that my being in *M* is partly or wholly constituted by my introspectively believing that I am in *M*, and (iii) my believing that I am in *M* and my being in *M* are not partly or wholly constituted by a third state of affairs.

---

[27] The notion of mereological distinctness discussed in Stoljar (2008, 266) does not sufficiently distinguish between spatio-temporal distinctness and constitutive distinctness. Both are a kind of mereological distinctness, having to do with parts and wholes, but they seem otherwise to be quite different relations.

As this definition reveals, constitutive distinctness has a similar logic to other mereological notions. In particular, constitutive distinctness allows for a distinction between whole and partial distinctness. Indeed, as far as states of affairs go, it is easy to make sense of the notions of being wholly or partly distinct.[28] A state of affairs can be wholly constituted by another, as in the case of the chair's being red and the chair's being coloured. A state of affairs can be partly constituted by another, as in the case of John's being a bachelor and John's being a man.

We can now cast the argument from distinctness in terms of constitutive distinctness:

The Argument from Constitutive Distinctness

P1. If my introspectively believing that I am in *M* is constitutively distinct from my being in *M*, then it is logically possible for me to introspectively believe that I am in *M* without being in *M*.

P2. My introspectively believing that am in *M* is constitutively distinct from my being in *M*.

C1. It is logically possible for me to introspectively believe that I am in *M* without being in *M*.

Put this way, P1 is plausibly true. Whenever two states of affairs are constitutively distinct, it is logically possible for one to obtain and for the other not to obtain and vice versa. None of the standard counter-examples touch this notion of distinctness, since all such examples involve things which are not constitutively distinct. Consider the example of the chair's being coloured and the chair's being red, or my knowing that it is raining and my believing that it is raining, or my intending to get wet and my wanting to get wet. None provide counter-examples because none involve states of affairs which are constitutively distinct but necessarily connected.

I can do little more to defend P1 here. It is prima facie plausible, and there are no obvious counter-examples to it. Yet it is hard to imagine what a positive argument for it would look like. And perhaps this explains why Armstrong offers no arguments in favour of it. But at least it is not obviously false as P1 of the argument from numerical distinctness is. And that will do for now. Moreover, something like Hume's dictum, understood in terms of constitutive distinctness, is often taken to be true elsewhere. In his discussion of phenomenal beliefs, David Chalmers writes: 'In the case of direct phenomenal beliefs, beliefs and experiences are *not* entirely distinct existences. It is precisely because of the constitutive connection between experiential quality and belief that the two can be necessarily connected' (Chalmers 2003, 224). The implication here seems to be that if there wasn't a constitutive relation between either, there couldn't be a necessary connection.

---

[28] See Sanford (1984, 76–8).

The question now is whether the distinctness premise is true: is my being in *M* constitutively distinct from my introspectively believing that I am in *M*? What reasons do we have for thinking so? The argument we gave in the case of numerical distinctness, which appealed to the numerical distinctness of the properties or states involved, cannot be used here, since we have seen that states of affairs involving numerically distinct states and properties can fail to be constitutively distinct.

For reasons which will become apparent, I do not think that there is any a priori argument available for the distinctness premise, and, moreover, I think that there may well be empirical arguments available which show that the premise is false. But before explaining why I think this in more detail, we will examine Armstrong's arguments for the distinctness premise with the question of what he must have meant by 'distinct' in mind.

3.3 Armstrong's Arguments for the Distinctness Premise

Armstrong's arguments for the distinctness premise in both IKI and *MTM* are rather brief. We will examine each in turn, and attempt to reconstruct a cogent argument out of his remarks. Here is the argument Armstrong offers in IKI in its entirety:

> [T]he apprehension of something must be distinct from the thing apprehended. For if not, we are faced with a flagrant circularity. Having a pain logically involves apprehension of—what? The pain itself! This is as bad as saying that to be a cat logically involves being the offspring of cats. It seems, therefore, that there must always be a distinction between *being* in a certain mental state and *being aware* that we are in that state. (Armstrong 1963, 423)

Without yet saying anything about what Armstrong means by 'distinct' here, we can see that he clearly takes non-distinctness to entail logical involvement, and he takes non-circularity to be a constraint on logical involvement. Eschewing Armstrong's talk of 'apprehension' for talk of introspective belief, we might reconstruct the argument as follows:[29]

---

[29] It is important for the moment that the argument be stated in terms of a relation of awareness to my being in *M*, rather than in terms of my introspectively believing that I am in *M*. The latter, but not the former, creates an intensional context with respect to *M*. See Jackson on this point:

The Argument for Distinctness (in IKI)

P1. If my believing that I am in *M* is not distinct from my being in *M*, then my being in *M* logically involves my introspectively believing that I am in *M*.

P2. It is not the case that my being in *M* logically involves my introspectively believing that I am in *M*.

C1. My introspectively believing that I am in *M* is distinct from my being in *M*.

What could Armstrong mean by 'logically involves' here? Whatever he means by it, clearly non-circularity is a constraint on logical involvement. It is plausible that by 'logically involves' Armstrong has something like analytic definition in mind, since non-circularity is a constraint on analytic definition. So, we might say that being a bachelor logically involves being a man and being unmarried: there exists an analytic definition of being a bachelor with the following form: to be a bachelor is to be an unmarried man.[30] So, let's say that one thing logically involves another if and only if the former has an analytic definition in terms of the latter.

With this understanding of logical involvement in mind, we can see that P2 is very plausible. My being in *M* does not logically involve my introspectively believing that I am in *M*. Being in *M* doesn't have an analytic definition in terms of believing that I am in *M* nor vice versa. It is not, however, that a definition of one in terms of the other would be circular, since in a definition of being in *M* in terms of introspectively believing that I am in *M*, *M* occurs in an intensional context (Armstrong's presentation in terms of apprehension obscures this and makes the claim of circularity more worrying). Let us grant that P2 is true then.

Now let's consider P1. Why does Armstrong think that it follows from the claim that my being in *M* and my being introspectively aware of my being in *M* are not distinct that my being in *M* logically involves my being introspectively aware of my being in *M*? After all, it doesn't follow from the claim that my being in *M* is not distinct from my being in *M* that my being in *M* logically involves my being in *M*. If it did, then Armstrong's argument would show that my being in *M* is distinct from my being in *M*, since the claim that my being in *M* logically involves my being in *M* is flagrantly circular! But perhaps Armstrong has a sense of 'distinct' in mind on which my being in *M* is not automatically not distinct from my being in

---

'Armstrong talks here of apprehension and awareness; but it is, as Armstrong himself notes, impossible to apprehend or be aware of a non-existent state, because "apprehend" and "aware" are "success verbs". So we will stick to our approach to incorrigibility which regards the central issue as whether mistaken *belief* is possible' (Jackson 1973, 57).

Here is what Armstrong says: '"Apprehension", of course, is a success-word, and strictly it would be out of place here. But it would be possible for there to be something just like the apprehension, except that the object of the apprehension did not exist' (Armstrong 1963, 422–3).

[30] Of course, the form of words 'to be a bachelor is to be an unmarried man' isn't a definition on its own any more than 'something is a bachelor if and only if it is an unmarried man' is a definition on its own. We can use these forms of words to give a definition. By convention, the thing being defined comes first and the things in terms of which it is being defined come later.

M, as it is if 'distinct' is understood as 'numerically distinct'. Constitutive distinctness will do the trick, since it is not the case that my being in *M* is constituted by my being in *M*. Constitution is arguably irreflexive: nothing is constituted by itself. It may well be that if my being in *M* is not constitutively distinct from my being in *M*, then my introspectively believing that I am in *M* logically involves (in the sense of analytic definition) my being in *M*. And there may have been a time when claims like these were taken to be plausible. But it is now widely held that there can be constitution without analytic definition. The chair's being coloured is constituted by its being red, but it isn't obvious at all that its being coloured logically involves its being red, that is, that its being coloured has an analytic definition partly in terms of its being red (disjunctive definitions of 'is coloured' in terms of 'is red, green, blue, or . . . ' being rather unfashionable).

So much for the argument of IKI then. Let us now turn to the argument of *MTM* for the distinctness premise. Armstrong's defence of the distinctness premise in *MTM* appears in the context of a response to the argument from infallibility against the central-state theory. Armstrong aims to give an argument against infallibility which is independent of the central-state theory.[31] Importantly, as we will see, Armstrong's argument is not independent of the traditional theory of introspection. It will be important to keep this in mind as we proceed. Moreover, Armstrong prefaces his argument for the distinctness premise with the concession that 'at a critical point in the argument I have been unable to *demonstrate* that what I say is necessarily true, although I believe it to be necessarily true, and offer some arguments for its being necessarily true' (*MTM*, 106). We will come to this lacuna in his argument in a moment. But first, the argument. Here is the first part:

> [L]et us consider the mechanical analogue of awareness of our own mental states: the scanning by a mechanism of its own internal states. It is clear here that the operation of scanning and the situation scanned must be 'distinct existences'. A machine can scan itself only in the sense that a man can eat himself. There must remain an absolute distinction between the eater and the eaten: mouth and hand, say. Equally, there must be an absolute distinction between the scanner and the scanned. Consider an eye (taken solely as a mechanism) scanning itself by means of a mirror. Certain features of the eye, such as its colour and shape, will register *on* the eye. But the registering will have to be something logically distinct from the features that are registered. (*MTM*, 106–7)[32]

So far, so good. Here is the second part:

---

[31] See (*MTM*, 103).

[32] This, by the way, is the context for Armstrong's only use of 'logically distinct'.

> Now what reason is there to think that awareness of its own states in the case, of, say, a spiritual substance, will differ in its logical structure from that of a self-scanning device in a mechanism? Why should the substitution of spiritual for material substance abolish the need for a distinction between object and subject?
> (*MTM*, 107)

Now, on a first reading, it is hard to discern the structure of Armstrong's argument, and the appeal to 'the mechanical analogue of awareness of our own mental states' can seem to be blatantly question-begging. To charitably reconstruct Armstrong's argument here, we need to use some imagination. I suggest that the overall structure of Armstrong argument is given by the following reconstruction, although Armstrong's focus in the text is on arguing for the premises of this argument (material in parentheses is material which is presupposed and not explicitly stated in the argument).

The Argument for Distinctness (in *MTM*)

P1. If the central-state theory is true, (and the traditional theory of introspection is true,) then my introspectively believing that I am in *M* is distinct from my being in *M*.

P2. If the central-state theory is false, (and the traditional theory of introspection is true,) then my introspectively believing that I am in *M* is distinct from my being in *M*.

C1. (If the traditional theory of introspection is true,) my introspectively believing that I am in *M* is distinct from my being in *M*.

With a little imagination we can see that this is the implicit structure of the overall argument. The argument for P1, also largely implicit, is the following. If the central-state theory is true, and the traditional theory of introspection is true, then our awareness of our own mental states is exactly like the scanning by a mechanism of its own internal states. For someone committed to the traditional theory of introspection and the central-state theory, the analogy with the scanning by a mechanism of its own internal states, is a close one. The claim about the primary subject, namely, scanning by a mechanism of its own internal states is that the states scanned and the states which are the result of the scanning process must be distinct.

The argument for P2 has the same structure as the argument for P1. If the central-state theory is false, and the traditional theory of introspection is true, then our awareness of our own mental states is like the scanning by a mechanism of its own internal states. Again, the claim about the primary subject is that the states scanned and the states which are the result of the scanning process must be distinct. Clearly, Armstrong thinks that the analogy is a good one, but all he can offer are the rhetorical questions already quoted above: 'what reason is there to think that awareness of its own states in the case of, say, a spiritual substance, will differ in its logical structure from that of a self-scanning device in a mechanism?'. He writes: 'I must admit that I can see no way to prove that there must be such a parallelism, which is a *lacuna* in my argument' (*MTM*, 107). So, while Armstrong does not have an argument for P2, he does not see any reason for thinking that it is not the case.

So much then for the overall structure of Armstrong's argument here and the structure of his argument for the premises. How plausible is his claim about the primary subject, namely, that in the case of the scanning by a mechanism of its own states, the states scanned and the states which are a result of that scanning are distinct? We may feel that they must be distinct. But Armstrong's remarks about the distinction between the eater and the eaten probably don't help. It is true that there is a sense in which a mechanism can scan itself but that there is also a sense in which that which does the scanning and that which gets scanned must be distinct. When a man eats himself, there is a part of him that does the eating and a part of him which gets eaten, and these are always distinct. Similarly, when a mechanism scans itself, there is a part of the mechanism that does the scanning and a part of the mechanism which gets scanned.[33] So, it is the parts which do the scanning and the parts which get scanned which must be distinct, according to Armstrong. It is here that 'there must be an absolute distinction between the scanner and the scanned' (*MTM*, 107). We may well be willing to grant Armstrong this claim about the logical structure of scanning, and so grant him the claim about the primary subject. But at best this supports a claim about numerical distinctness and not constitutive distinctness.[34] Why couldn't the scanner and the scanned fail to be constitutively distinct?

So much for Armstrong's arguments for the distinctness premise then. Neither seems to offer independent support for the claim that my introspectively believing that I am in *M* is constitutively distinct from my being in *M*. The final question I want to take up in this section is whether Armstrong meant 'numerically distinct' or 'constitutively distinct' by 'distinct'. There are reasons, beyond reasons of charity, for thinking that he had something like the latter in mind. The first reason for thinking that Armstrong doesn't mean 'numerically distinct' by 'distinct' is that he provides an example of a necessary connection between numerically distinct existences which are nonetheless not distinct existences. In *MTM*, before offering the argument we have just been discussing, Armstrong writes that

---

[33] Armstrong clearly thought there was a parallel in the case of introspection. Notice the appeal to parts in the following passage. This way of talking is implicit in the following from Armstrong: 'consciousness of our own mental state becomes simply the scanning of one part of our central nervous system by another. Consciousness is a self-scanning mechanism in the central nervous system' (Armstrong 1980a, 15).

[34] See Jackson (1973).

'X's shape is not a "distinct existence" from X's size, in the way that X's colour and X's smell could be said to be "distinct existences"' (*MTM*, 106). However, if something has a shape, it has a size. Armstrong is willing to allow that 'the incorrigibilist can claim, awareness of a pain stands to the pain as a thing's shape stands to its size, or its colour stands to its extension' (*MTM*, 106). That is, the incorrigibilist (the infallibilist) can claim that awareness of a pain and pain are not distinct existences in the sense that something's shape and its size are not distinct existences (even though something's shape and something's size are not numerically identical). This is evidence that by 'distinct' Armstrong did not mean 'numerically distinct'. And the second reason is that Armstrong happily uses the modifiers 'partial' and 'wholly' in connection with identity and distinctness.[35] Since numerical identity does not admit of such modifiers, it is doubtful that Armstrong could have had it in mind.

3.4 Sanford and Armstrong on Distinctness

I left the discussion of the distinctness premise hanging on a sceptical note, and we have just seen that Armstrong's arguments do little to establish the premise, understood in terms of constitutive distinctness. It turns out that in later work Armstrong all but concedes that he has no good argument for this premise. It will be instructive for us at this stage to examine his concession. It occurs in an exchange between Armstrong and David Sanford. In 'Armstrong's Theory of Perception', Sanford writes: 'Armstrong's argument for the incompatibility of central-state materialism and the incorrigibility of introspective knowledge overlooks the possibility that the scanning process might have the item scanned as a proper part' (1984, 78).

Sanford casts his objection in terms of mereological distinctness, but we can charitably take it to concern constitutive distinctness. Sanford's concern is that Armstrong has overlooked the possibility that while my believing that I am in *M* and my being in *M* are not identical, they are not wholly distinct either. Sanford writes: '. . . if two things, states, or processes, are not identical, it does not follow that they are wholly distinct' (1984, 77). The possibility Sanford is concerned with has been overlooked, and can be summed up as follows. Let M^* refer to the state of believing that I am in *M*. Let *B* and B^* refer to state of the central nervous system. Suppose $M = B$ and $M^* = B^*$, as the central-state theory holds. How do we know that B and B^* are wholly distinct? Is this not a matter for neurophysiological research? And if they are not wholly distinct, then my being in M and my being in M^* will not be wholly distinct. In other words, it looks as though the question of distinctness is an empirical one.

---

[35] See Armstrong (1984b).

Armstrong concedes the point in his reply to Sanford:

> [I]n arguing for the corrigibility of introspective awareness in *A Materialist Theory*, I assumed that the state introspected was a wholly distinct existence from the introspective awareness of it. But, Sanford argues, the mental processes which constitute, say, pain and the awareness of pain, might be partly identical . . . My argument was therefore incomplete . . . Nor do I know how to make it complete. (Armstrong 1984b, 232)

Nonetheless, Armstrong concludes his reply by raising some doubts about whether Sanford has shown that the distinctness premise is false:

> [I] do not think that Sanford has actually *falsified* my assumption that . . . in introspective awareness . . . [the] awareness [is] wholly distinct from . . . the mental state we are aware of. Sanford himself does not contest the point (private communication). He has shown that my argument contains a gap. He has not shown that it cannot be filled. On the other hand, I do not know how to fill it. Sanford has suggested to me, and I accept his suggestion, that future neurophysiological research may again be what is needed to cast light upon the issue.
> (Armstrong 1984b, 233)

So, Armstrong seems to leave the truth of the distinctness premise open to empirical hostage. Future neurophysiological research may tell us whether my introspectively believing that I am in M is constitutively distinct from my being in M. Whatever kind of research is needed, I think that this leads to an important point in the debate over distinctness: it is a contingent, empirical, matter whether my believing that I am in M is constitutively distinct from my being in M. Or, at least, I will argue that it is a contingent matter if we accept a broadly functionalist approach to mental states.

To make this point, we need only reflect on the coherence of the hypothesis that my introspectively believing that I am in M might not be numerically distinct from my being in M on a broadly functionalist approach to mental states. Let M^* refer to the state of believing that I am in M. Suppose that the state of the central nervous system which happens to play the M-role is B and the state which plays the M^*-role is B^*. It might turn out that $B = B^*$. It might not. If it did, then it would follow, at least on the identity theory, that $M = M^*$. It is a contingent matter whether $M = B$ and whether $M^* = B^*$. So, it is a contingent matter whether $M = M^*$. Indeed, it seems that the only way to directly exclude this possibility on a broadly functionalist approach to mental states would be to build some kind of distinctness assumption into the statement of the relevant roles.[36] In Lewisian terms, we can rule out this possibility from the start by insisting that the states involved in the Ramsey sentence are numerically distinct.[37] However, it would be a bad idea for the functionalist to build this kind of restriction into the statement of the roles, for surely if it turned out that the state which plays the *M*-role is identical to the state which plays the *M**-role, we would not conclude that we are not in *M* or *M** when we are in the state which plays both roles.

Now let us consider also the coherence of the hypothesis that my believing that I am in *M* might not be constitutively distinct from my being in *M*. Suppose that the state which plays the *M*-role is *B* and the state which plays the *M**-role is *B**. It might turn out that *B* is not constitutively distinct from *B**. It might not. If it did, then it would follow, at least on the identity theory, that *M* is not constitutively distinct from *M**. It is a contingent matter whether $M = B$ and whether $M^* = B^*$. So, it is a contingent matter whether *M* is not constitutively distinct from *M**. In Lewisian terms, we can rule out this possibility from the start by insisting that the states involved in the Ramsey sentence are constitutively distinct.[38] Again, if one thought that the folk theory really left the distinctness question open, then one would have to admit the possibility of discovering that what seem to be distinct mental states are not in fact distinct. And, again, it would be unwise to rule out the possibility from the start.

Because Armstrong identifies mental states with states of the central nervous system which play certain roles, he opens himself to the entailments we noted above concerning the non-distinctness of the physical states which play the relevant roles. In contrast, someone who identifies mental states with states of being in some state or other which play certain roles avoids these entailments. They can hold that *M* and *M** are distinct states, even if the states which play the *M*-role and the *M**-role are not distinct. It could well be that *B* and *B** play the *M*-role and the *M**-role respectively, and that *B* and *B** are numerically identical. Nothing would follow concerning the numerical identity or numerical non-identity of *M* and *M**. This would be a case, however, where my being in *M* and my being in *M** would not be distinct, since my being in *M* would be partly constituted by my being in *B* and my being in *M** would also be partly constituted by my being in *M*. And it could well be that *B* and *B** play the *M*-role and the *M**-roles respectively and that *B* and *B** are not constitutively distinct. This would also be a case where my being in *M* and my being in *M** would not be constitutively distinct. These options parallel the options above, however, the functionalist can hold that my being in *M* and my being in *M** can fail to be distinct if *B* and *B** fail to be distinct, whereas the identity theorist cannot, it seems, allow for this possibility.

---

[36] Of course, it may turn out that the roles are such that no single state could play both roles, but I see no grounds for this in the case of *M** and *M*.

[37] See Lewis (1983, 81).

[38] See Sanford (2005) on the misuse of the identity predicate to translate claims about distinctness in logic textbooks. Of course, again, it may turn out that the roles are such that no non-distinct states could play both roles, but I see no grounds for this in the case of *M** and *M*.

So where are we left when it comes to the argument from distinctness? I have suggested that P1, the no-necessary connections premise of the argument, is plausibly true when distinctness is understood as constitutive distinctness, but we have no good reason for thinking that P2, the distinctness premise, is true. So, the argument fails. And Armstrong admits as much. We have just seen that, at least from a broadly functionalist perspective, it is an empirical question whether the states which play the *M*-role and the *M**-role are numerically or constitutively distinct, and this has a direct bearing on whether *M* and *M** are constitutively distinct.

This brings us to the end of our discussion of Armstrong's argument from distinctness. We have seen that the argument is ultimately inconclusive. Understood in terms of numerical distinctness, it has an obviously false premise. Understood in terms of constitutive distinctness, a central premise of the argument is left hanging. As we will see in Section 4 when we turn to Sydney Shoemaker's arguments against the self-scanning theory and inner-sense theories in general, it is exactly the assumption of distinctness that Shoemaker denies.

4. The Argument from Non-Distinctness

Anyone familiar with the work of Sydney Shoemaker will know that he is meant to have an argument against the self-scanning theory and inner-sense theories more generally. And you will also know that the argument has something to do with the logical or metaphysical impossibility of self-blindness. Indeed, upon reading Shoemaker, you will come away with this impression, but may not know exactly what the argument against the self-scanning theory is meant to be, nor know exactly what self-blindness is, nor know exactly what Shoemaker's argument for the logical or metaphysical impossibility of self-blindness is. The aim of this section is to see whether we can reconstruct an argument from Shoemaker's discussions which threatens the self-scanning theory. I will argue that we can, but only if we understand Shoemaker's argument for one of the key premises of the argument from non-distinctness as an empirical argument to the best explanation, and not, as Shoemaker seems to understand it, as an a priori deduction of the non-distinctness of introspective beliefs and the mental states they are about.

We begin with Shoemaker's understanding of self-blindness. Shoemaker defines self-blindness with respect to mental state types as follows.

> **Self-Blindness**: to be self-blind with respect to some mental state type is to have the ability to be in states of that type, to have the concept of oneself and of being

> in particular states of that type, to be sufficiently intelligent and sufficiently rational, and yet to be unable to determine by introspection alone whether one is in a particular mental state of that type.[39]

The main obscurity in Shoemaker's presentation concerns the appeal to sufficient intelligence and sufficient rationality. Shoemaker nowhere carefully defines these. But it isn't clear that he needs to. Let's work with the following operational understanding: whatever 'sufficient intelligence' and 'sufficient rationality' mean, anyone reading this chapter qualifies as sufficiently intelligent and sufficiently rational. Now, what Shoemaker says about self-blindness is that it is logically, conceptually, or metaphysically impossible. Since anyone reading this chapter will have the ability to believe various propositions, what this amounts to is the claim that any minimal duplicate with the same conceptual capacities, intelligence, and rationality, will be able to determine by introspection alone whether they believe some proposition or not. Shoemaker sometimes puts his claim in terms of supervenience: there can be no difference in facts about whether one has an ability to determine by introspection alone whether one is in a particular mental state of certain type without a difference in the facts about conceptual capacities, rationality, intelligence, or the ability to be in states of that type.

A cursory reading of Shoemaker suggests that his case against the self-scanning theory can be captured by what I will call the argument from the impossibility of self-blindness:

The Argument from the Impossibility of Self-Blindness

P1. If the self-scanning theory of introspection is true, then self-blindness is metaphysically possible.

P2. Self-blindness is not metaphysically possible.

C1. The self-scanning theory of introspection is not true.

There is a lot we could say about P2, but I am willing to grant that, whatever Shoemaker's arguments for P2 are, P2 is plausibly true.[40] (In any case, it won't matter for our understanding of the argument from distinctness below.) I want to focus instead on P1. Why does Shoemaker think that P1 is true? In characterizing the kind of view he is opposed to, he writes:

---

[39] I take this to be the essence of Shoemaker's claims about self-blindness. In 'On Knowing One's Own Mind', where he first introduces the idea of self-blindness, Shoemaker writes:

> 'What I wish to maintain is the impossibility of something I shall call "self-blindness". A self-blind creature would be one which has the conception of the various mental states, and can entertain the thought that it has this or that belief, desire, intention, etc., but which is unable to become aware of the truth of such a thought excpepct in a third-person way.'
>
> (Shoemaker 1996a, 30–1)

There is no mention here of sufficient intelligence and rationality. Shoemaker emphasizes these elements in later work. See esp. Shoemaker (1996c, 236).

[40] See Kind (2003, 45) for some dissent.

> it will be held that for each kind of mental fact to which we have introspective access, it is at least logically possible that there should be creatures in which such facts obtain, and who have the ability to conceive of them, but who are self-blind with respect to them. This I take to be a consequence of the independence condition that is built into the broad perceptual model.
> (Shoemaker 1996c, 226)

And:

> the existence of mental entities and mental facts is, logically speaking, as independent of our knowing about them introspectively, and of there being whatever means or mechanisms enable us to know about them introspectively, as the existence of physical entities and physical facts is of our knowing about them perceptually, or of there being the means or mechanisms that enable us to have perceptual knowledge of them. This is the view that implies that our introspective self-knowledge should be construed in terms of what in my first lecture I called the broad perceptual model. The core of the stereotype of perception underlying that model consisted of two conditions, one of them (call it the *causal condition*) saying that our beliefs about our mental states are caused by those mental states, via a reliable belief-producing mechanism, thereby qualifying as knowledge of those states and events, and the other (call it the *independence condition*) saying that the existence of these states and events is independent of their being known in this way, and even of there existing the mechanisms that make such knowledge possible. (Shoemaker 1996c, 224–5)

This can initially seem rather curious. Don't we all agree that someone can be in a mental state without introspectively believing that they are in that mental state or without having the ability to determine by introspection alone whether they are in that mental state? Even by Shoemaker's own lights, this is possible in those who lack the relevant conceptual capacities, a sufficient level of rationality, or a sufficient level of intelligence. The crucial clue here is the expression 'as independent'. In the case of sense-perception, it is possible for someone to have the relevant conceptual capacities, be sufficiently rational and intelligent, and yet lack the ability to determine by a particular perceptual modality whether some external thing is in some state. In characterizing the independence condition, Shoemaker clearly *wants* the possibility of self-blindness to follow, and he thinks that the analogy with perception establishes it. In an earlier presentation Shoemaker writes: 'Only if self-blindness were a conceptual possibility would it be appropriate to think of the capacity for self-acquaintance as a quasi-perceptual capacity, which is something over and above the capacity to have and conceive of the mental states in question' (1996a, 31).

Now, one might read someone like Armstrong as being committed to the idea that the capacity for self-scanning is something 'over and above' having the ability to be in a mental state of a particular kind, being sufficiently rational and intelligent, and having the relevant conceptual capacities. One might get Armstrong to agree perhaps that some of our ancestors had the same levels of rationality, intelligence, and conceptual capacities, but hadn't yet fully evolved an ability to determine by introspection alone whether they were in particular mental states of that type. But then again, one might not, since Armstrong thinks that introspection is necessary for a sufficient level of rationality. So, it isn't even clear that Armstrong will think of the capacity for introspection as being something 'over and above' these other capacities. In any case, the self-scanning theory holds that introspective beliefs are arrived at by a self-scanning process, and it can take or leave the additional claim that the capacity for self-scanning is something over and above these other capacities. After all, the inner-sense theorist isn't committed to introspection being like perception in every sense.[41]

There seems to be a growing consensus in the self-knowledge literature that if the argument from the impossibility of self-blindness is Shoemaker's argument, then it is ultimately unsuccessful: even granting the truth of the impossibility of self-blindness premise, the argument does nothing to undermine the self-scanning theory. Brie Gertler writes of Shoemaker's argument that: 'The main problem is that these objections center on the idea that self-awareness *will* be achieved by rational creatures. But they do not directly concern the crux of the inner sense theory, which concerns *how* self-awareness is achieved' (Gertler 2011, 158). I agree: *if* the argument from the impossibility of self-blindness is all Shoemaker has against Armstrong, Armstrong needn't be worried. But I don't think that this is all Shoemaker has. To stop here is to totally ignore other aspects of Shoemaker's arguments. It is to ignore the functionalist elements and the constitutivist elements. I want to suggest that the best way of understanding Shoemaker's arguments against the self-scanning theory is in terms of the following argument.

The Argument from Constitutive Non-Distinctness

P1. If the self-scanning theory of introspection is true, then my introspectively believing that I am in *M* is constitutively distinct from my being in *M*.

P2. My introspectively believing that I am in *M* is not constitutively distinct from my being in *M*.

C1. The self-scanning theory of introspection is not true.

---

[41] See Lycan (1987, 28).

One merit of this argument is that P1, as we have seen, is plausibly a commitment of the self-scanning theory. As we have seen, Armstrong is willing to admit that P2 might be true. We can now see various elements of Shoemaker's discussion as attempting to establish P2. This is the constitutivist element of Shoemaker's argument. Here comes the functionalist element. It gives us the argument for constitutive non-distinctness:

The Argument for Constitutive Non-Distinctness

P1. If the state which plays the *M*-role is numerically identical to the state which plays the *M**-role, then my being in *M* is not constitutively distinct from my being in *M**.

P2. The state which plays the *M*-role is numerically identical to the state which plays the *M**-role.

C1. My being in *M** is not constitutively distinct from my being in *M*.

Notice that we have introduced numerical identity back into our discussion, even though we are ultimately concerned with constitutive distinctness. We already saw how the two are related from a functionalist perspective, so P1 of this argument should be relatively uncontroversial. So if P2 is true, the conclusion follows. And the conclusion is P2 of the argument from constitutive non-distinctness. So if P2 is true, the conclusion of the argument from constitutive non-distinctness goes through and the self-scanning theory of introspection is in trouble. Here is what I take to be a crucial statement of this argument from Shoemaker:

> If, as I have suggested, believing that one believes that *P* can be just believing that *P* plus having a certain level of rationality, intelligence, and so on, so that the first-order belief and the second-order belief have the same core realization, then it will be altogether wrong to think of the second-order belief in such cases as *caused* by the first-order belief it is about. Here the relation of an 'introspective' belief to the state of affairs it is about is altogether different from the relation of a perceptual belief to the state of affairs it is about... [O]ur introspective access to our own mental states differs markedly from our perceptual access to things in our environment. (Shoemaker 1996c, 244)

Shoemaker's reference to the first-order belief and the second-order belief having the same core realization is equivalent to our claim that the state which plays the *M*-role is numerically identical to the state which plays the *M**-role. To overlook this is to overlook a crucial element in Shoemaker's argument against the self-scanning theory. And, as I said, this argument has the advantage of appealing to an uncontroversial commitment of the self-scanning theory. So, everything comes down to Shoemaker's arguments for the numerical identity premise of the argument for constitutive non-distinctness.

As we established earlier, we have two coherent hypotheses: either the state which plays the *M*-role is numerically distinct from the state which plays the *M**-role or it is not. Call the latter the identity hypothesis, and the former the non-identity hypothesis. It is a contingent matter which hypothesis is true so we can hardly expect an a priori argument here (unless we think that the truth of one hypothesis is contingent and a priori!). We could await 'future neurophysiological research'. But Shoemaker thinks we needn't. We've got a claim about supervenience, which needs explaining (and here it doesn't matter what strength the supervenience claim has) and, according to Shoemaker, the best explanation is that the state which plays the *M*-role is numerically identical to the state which plays the *M**-role.

Let's consider the two hypotheses in more detail to see why this should be so. According to the non-identity hypothesis, adding rationality, intelligence, and conceptual competence to the ability to be in mental states of a particular type ensures that there is a law-like connection between being in a state which plays the *M*-role and being in a numerically distinct state which plays the *M**-role (or perhaps even a constitutively distinct state). Here is how Shoemaker characterizes this hypothesis:

> A natural way of thinking is to say that the belief that *P* is one thing, the introspective belief that one believes that *P* is another, and that introspective awareness consists in the first giving rise to the second. Going with this is a way of thinking about how those mental states are realized in the brain. The belief that *P* is realized in one neural state, perhaps a highly distributed one, the belief that one believes that *P* is realized in another, and the brain is so wired that, under certain condition, the first neural state gives rise to the second.
>
> (Shoemaker 1996c, 242)

According to the identity hypothesis, adding rationality, intelligence, and conceptual competence to the ability to be in mental states of a particular type ensures that the state which plays the *M*-role also plays the *M**-role. Here is how Shoemaker characterizes this hypothesis:

> all you have to add to the available first-order belief, in order to get the second-order belief is the appropriate degree of intelligence, etc. It is not that adding this pushes the creature into a new state, distinct from any it was in before, which is the core realization of believing that it has this belief. It is rather that adding this enables the core realization of the first-order belief to play a more encompassing causal role, one that makes it the core-realization of the second-order belief as well as the core-realization of the first-order belief. (Shoemaker 1996c, 244)

Now, the supervenience claim, suitably weakened, would be satisfied if rationality, intelligence, and conceptual competence ensured either that there was a law-like connection between numerically distinct states which play the relevant roles or that a single state played both roles. The question now is why think that the identity hypothesis provides the best explanation? Does Shoemaker have an argument? My view is that Shoemaker never presents an argument that one explanation is better than the other. Rather he just puts the two hypotheses on the table and judges that the latter provides a better explanation and hopes that we will too. He is strikingly non-committal in his presentations of the hypotheses. In 'Knowing One's Own Mind', he writes:

> But perhaps we needn't suppose that each belief or desire produces a separate state which is the knowledge or belief in its own existence. Perhaps instead it is the case that insofar as a person is rational, each belief and desire tends to double as knowledge or belief in its own existence. For perhaps in making its contribution to the preservation or restoration of consistency and coherence in the system, each desire and belief plays the causal role appropriate to a second-order belief to the effect that one has the first-order belief or desire in question. (Shoemaker 1996a, 33–4)

Perhaps, perhaps, perhaps, indeed. But why? Shoemaker is no more confident in 'Self-Knowledge and Inner-Sense II: The Broad Perceptual Model':

> it might be that [the first-order state and the belief about it] have the same core realization . . . I suggest that this is how it might be when the two states are, respectively, an available first-order belief, say that the sun is shining, and the second-order belief that one has the first-order belief, say the belief that one believes that the sun is shining. (Shoemaker 1996c, 243)

Again, it might be. It might not. Why should we think it is? I have suggested, then, that nothing forces one hypothesis onto us rather than the other, no matter how appealing we might think it is. Shoemaker does sometimes speak as if one, rather than the other, is forced on us. But, as we have seen, this is implausible. It is an empirical matter whether the state which plays the M-role plays the M^*-role. Upon discovering that the state which plays the M-role does not play the M^*-role, we wouldn't conclude that we aren't rational, aren't intelligent, or lack the relevant conceptual capacities. We'd just conclude that the non-identity hypothesis is correct. So, the best way of understanding Shoemaker's argument here is as an argument to the best explanation. And understood this way, the argument is quite plausible.

Indeed, in some of Shoemaker's more recent work there is an auxiliary argument for the conclusion that the identity hypothesis provides the best explanation.[42] Shoemaker notes that not only is there a positive law-like connection between being in *M* and being in *M** if you are rational, intelligent and have the relevant conceptual capacities (that is, if you come to be in *M*, you will come to be in *M**), but there is also a negative law-like connection (if you cease to be in *M*, you will cease to be in *M**). The point is that introspective beliefs do not have a cognitive dynamics independent of the cognitive dynamics of the states they are about: they come and go together. Now in order to explain the negative connection on the numerical distinctness hypothesis, the scanner will need to be constantly scanning for the absence of mental states. Entering *M* must cause you to enter *M**, but ceasing to be in *M* must cause you to cease to be in *M**. Again, it isn't impossible to explain this on the numerical distinctness hypothesis, but the numerical identity hypothesis has a much better explanation, since it straightforwardly predicts that your being in *M** will not survive your ceasing to be in *M*.

So, there is significant plausibility to Shoemaker's claim that the identity hypothesis provides the best explanation of the connection between introspection, rationality, intelligence, and conceptual capacity. And so Shoemaker has a good argument for the non-distinctness premise. Given the plausibility of the other premise of the argument from non-distinctness, he has a good argument against the self-scanning theory.

5. Distinctions Without a Difference?

To sum up, I have argued that Armstrong's argument from distinctness against infallibility is ultimately unsuccessful, as he admits, because he is unable to establish the premise that my introspectively believing that I am in some mental state is distinct from my being in that state. We have examined Shoemaker's argument for the conclusion that my introspectively believing that I am in some mental state is *not* distinct from my being in that mental state and for the conclusion that the self-scanning theory is false. I argued that, understood as a lowly argument to the best explanation, Shoemaker's argument for this conclusion is plausible. So, the self-scanning theory, as Armstrong understands it, is in trouble. I want to wrap up the discussion with a few brief remarks about the depth—or perhaps indeed the shallowness—of the disagreement between Armstrong and Shoemaker on the matter of distinctness.

Consider the following passage from Shoemaker:

> How *can* we make, noninferentially, judgements that have complex causal implications? Here is a sketch of an answer. Our minds are so constituted, our

---

[42] See Shoemaker (2012).

> brains are so wired, that, for a wide range of mental states, one's being in a certain mental state produces in one, under certain conditions, the belief that one is in that mental state. This is what our introspective access to our own mental states consists in. The 'certain conditions' may include one's considering whether one is in that mental state, and will certainly include what is a precondition of this, namely that one has the concept of oneself and the concept of the mental state. The beliefs thus produced will count as knowledge, not because of the quantity or quality of the evidence on which they are based (for they are based on no evidence), but because of the reliability of the mechanism by which they were produced. And now we add that the concepts under which these states are conceived in these beliefs are causal or functional ones. (Shoemaker 1996b, 222)

Shoemaker goes on: 'The account just sketched is very close to what I believe. But it is also close to what has been maintained by philosophers—here I particularly have in mind David Armstrong—who are enthusiastic supporters of the idea that introspection is a kind of perception' (Shoemaker 1996b). What is the disagreement, then, between the two? It is largely over what Shoemaker calls the 'independence condition' (Shoemaker 1996b, 223). As Shoemaker writes: 'I think the fundamental difference between perception and introspection is the failure of the latter to satisfy the "independence condition"' (Shoemaker 1996b, 244). But if we understand this as a question concerning constitutive distinctness, then what is the big deal? How could the difference between constitutive and causal relations in introspection carry so much weight, and mightn't Armstrong just understand the self-scanning theory in a more inclusive way, playing down the importance of the distinction between constitution and causation? Mightn't Armstrong happily concede the point, since this kind of non-distinctness is perfectly compatible with introspective fallibility and so can't be used to motivate an argument against central-state materialism? Armstrong may well concede that introspection isn't exactly like perception on the traditional model of introspection. Indeed, Armstrong sometimes says that the important sense in which introspection is like perception is that it involves the 'flow of information' (Armstrong 1973, 22). Arguably, there would be a flow of information between mental states and our beliefs about them whether or not the relation between the states was causal or constitutive. In light of these considerations, it is hard to avoid the conclusion, looking back, that the disagreement between Armstrong and Shoemaker over distinctness isn't over a particularly deep matter, and the difference between Shoemaker and Armstrong on introspection really starts to look diminishingly small.[43]

---

[43] The two philosophers, of course, take quite different approaches to the *epistemology* of introspection. But it seems that they have more in common on the *metaphysics* of introspection than may first appear.

References

Armstrong, D. M. (1961) *Perception and the Physical World*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1962) *Bodily Sensations,* London: Routledge & Kegan Paul.

Armstrong, D. M. (1963) 'Is Introspective Knowledge Incorrigible?', *The Philosophical Review*, 72(4): 417–32.

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1973) *Belief, Truth and Knowledge*, Cambridge: Cambridge University Press.

Armstrong, D. M. (1976) 'Incorrigibility, Materialism and Causation: A Reply to George Pappas', *Philosophical Studies*, 30: 125–7.

Armstrong, D. M. (1980a) 'The Nature of Mind', in *The Nature of Mind and Other Essays,* St Lucia: University of Queensland Press, 1–15.

Armstrong, D. M. (1980b) 'What is Consciousness?', in *The Nature of Mind and Other Essays*, St Lucia: University of Queensland Press, 55–67.

Armstrong, D. M. (1984a) 'Consciousness and Causality', in *Consciousness and Causality: A Debate on the Nature of Mind*, by D. M. Armstrong and Norman Malcolm, Oxford: Basil Blackwell, 103–92.

Armstrong, D. M. (1984b) 'Replies', in *D. M. Armstrong, Profiles*, vol. 4, edited by Radu J. Bogdan, Dordrecht: D. Reidel, 224–72.

Armstrong, D. M. (1984c) 'Self-Profile', in *D. M. Armstrong, Profiles*, vol. 4, edited by Radu J. Bogdan, Dordrecht: D. Reidel, 3–54.

Armstrong, D. M. (1989) *A Combinatorial Theory of Possibility*, Cambridge: Cambridge University Press.

Armstrong, D. M. (1997) *A World of States of Affairs*, Cambridge: Cambridge University Press.

Baier, Kurt (1962) 'Smart on Sensations', *Australasian Journal of Philosophy*, 40(1): 57–68.

Carruthers, Peter (2011) *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford: Oxford University Press.

Chalmers, David J. (2003) 'The Content and Epistemology of Phenomenal Belief', in *Consciousness: New Philosophical Perspectives*, edited by Q. Smith and A. Jokic, Oxford: Oxford University Press, 220–72.

Dretske, Fred (2003) 'How Do You Know You Are Not a Zombie?', in *Privileged Access: Philosophical Accounts of Self-Knowledge*, edited by Brie Gertler, Burlington, MA: Ashgate Publishing, 1–14.

Dretske, Fred (2012) 'Awareness and Authority: Skeptical Doubts', in *Introspection and Consciousness*, edited by Declan Smithies and Daniel Stoljar, Oxford: Oxford University Press, 49–64.

Gertler, Brie (2011) *Self-Knowledge*, London: Routledge.

Jackson, Frank (1973) 'Is There a Good Argument Against the Incorrigibility Thesis?', *Australasian Journal of Philosophy*, 51(1): 51–62.

Kind, Amy (2003) 'Shoemaker, Self-Blindness and Moore's Paradox', *Philosophical Quarterly*, 53(210): 39–48.

Lewis, David (1972) 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy*, 50(3): 249–58.

Lewis, David (1983) 'How to Define Theoretical Terms', in *Philosophical Papers*, vol. 1, Oxford: Oxford University Press, 78–96.

Lycan, William G. (1987) *Consciousness*, Cambridge, MA: MIT Press.

Lycan, William G. (1996) *Consciousness and Experience*, Cambridge, MA: MIT Press.

Lyons, William E. (1986) *The Disappearance of Introspection*, Cambridge, MA: MIT Press.

Nichols, Shaun and Stephen P. Stich (2003) *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, Oxford: Oxford University Press.

Pappas, George S. (1976) 'Incorrigibility and Central-State Materialism', *Philosophical Studies*, 29: 445–56.

Sanford, David H. (1984) 'Armstrong's Theory of Perception', in *D. M. Armstrong, Profiles*, vol. 4, edited by Radu J. Bogdan, Dordrecht: D. Reidel, 55–78.

Sanford, David H. (2005) 'Distinctness and Non-Identity', *Analysis*, 65(4): 269–74.

Shoemaker, Sydney (1996a) 'On Knowing One's Own Mind', in *The First-Person Perspective and Other Essays,* Cambridge: Cambridge University Press, 3–24.

Shoemaker, Sydney (1996b) 'Self-Knowledge and "Inner-Sense", Lecture I: The Object Perception Model', in *The First-Person Perspective and Other Essays*, Cambridge: Cambridge University Press, 201–23.

Shoemaker, Sydney (1996c) 'Self-Knowledge and "Inner-Sense", Lecture II: The Broad Perceptual Model', in *The First-Person Perspective and Other Essays*, Cambridge: Cambridge University Press, 224–45.

Shoemaker, Sydney (2012) 'Self-Knowledge and Second-Order Belief', in *Introspection and Consciousness*, edited by Declan Smithies and Daniel Stoljar, Oxford: Oxford University Press, 239–58.

Smart, J. J. C. (1962) 'Brain Processes and Incorrigibility', *Australasian Journal of Philosophy*, 40(1): 68–70.

Stoljar, Daniel (2008) 'Distinctions in Distinction', in *Being Reduced: New Essays on Reduction, Explanation, and Causation*, edited by Jakob Hohwy and Jesper Kallestrup, Oxford: Oxford University Press, 264–78.

Williamson, Timothy (2000) *Knowledge and Its Limits*, Oxford: Oxford University Press.

Zimmerman, Aaron Z. (2008) 'Self-Knowledge: Rationalism vs. Empiricism', *Philosophy Compass*, 3(2): 325–52.

10

Armstrong's Just-so Story about Consciousness

Daniel Stoljar

1. Introduction

D. M. Armstrong's *A Materialist Theory of the Mind* (hereafter *MTM*) is a prime source of many ideas that are still widely discussed in contemporary philosophy of mind. Among these are:

- the causal or functionalist analysis of belief as a state apt to cause a certain sort of behaviour; and the correlative analysis of a purpose (i.e., an intention or desire) as an information-sensitive mental cause;
- the analysis of perception in terms of belief: perception as the getting of belief about one's immediate surroundings;
- the analysis of other mental states in terms of perception, and so ultimately in terms of belief: sensation as a sort of perception of one's own body, introspection as a sort of perception of one's own mind;
- the analysis of a conscious mental state as a state that is the target of a certain sort of introspection, or inner perception; and
- the distinctive two-premise argument for the identity of mental states with physical states of the central nervous system. The first premise of the argument, which follows from his causal analysis, is that mental states are states apt to produce a certain sort of behaviour. The second, empirical, premise is that C-fibers firing and other neurophysiological states are, in fact, states apt to produce that sort of behaviour. The conclusion is that mental states are physical states.

Perhaps less well known is Armstrong's claim that his book provides 'an answer to the question asked again and again by so many biologists and psychologists "What is the biological function of consciousness?"' (*MTM*, 162). In particular, he suggests at several points that his views about purpose and about introspection may be combined to provide a 'just-so story about introspective consciousness', as W. G. Lycan later described it (see Lycan 1996).

Daniel Stoljar, *Armstrong's Just-so Story about Consciousness* In: *Armstrong's Materialist Theory of Mind.* Edited by: Peter R. Anstey and David Braddon-Mitchell, Oxford University Press. © Daniel Stoljar 2022.
DOI: 10.1093/oso/9780192843722.003.0011

My topic here is this just-so story—Armstrong's account of the function of consciousness, as I will usually say. After providing an interpretation of the account (Sections 2 and 3), I will develop two critical points about it. The first (Section 4) is that appreciating this aspect of Armstrong's view blurs the difference between his own perceptual model of introspection, and the anti-perceptual models advanced by such critics as Sydney Shoemaker. The second (Section 5) is that the account is implausible since it conflates two interpretations of its central idea, namely, that 'a purpose is an information-sensitive mental cause'. In Section 6 I will bring these two criticisms together by noting a connection between Armstrong's position and an apparently different argument by Shoemaker; both fail, I will argue, for the same reason. Section 7 is a brief conclusion.[1]

2. Armstrong's Account

Armstrong's account is contained in two main chapters of *MTM*, Chapter 15 and Chapter 6. Chapter 15 summarizes his overall attitude to introspection, which he says is crucial for the argument in the preceding chapters. The summary concludes with this passage:

> If there are to be purposive trains of mental activity, then there must equally be some means by which we become apprised of our current mental state. Only so can we adjust mental behaviour to mental circumstances. For instance, if we are doing a calculation 'in our head' we will need to become aware of the current stage in the mental calculation that we have reached. Only if we do become so aware will we know what to do next. So there must be a way of becoming aware of our current mental state, which means that there must be introspection.
>
> (*MTM*, 327)

Similar passages occur earlier, in Chapter 6, which sets out his account of purposes:

> If the train of mental states between stimulus and response is to be a *purposive* affair . . . there [is] some method whereby information about the current state of affairs can be fed back to the driving force behind the teleological sequence. But if the teleological sequence in question is a sequence of events in the mind, this means that the agent must become informed of what is currently going on in his

---

[1] It goes without saying that the contemporary literature relevant to the evolutionary function of consciousness is much bigger and more sophisticated now than when Armstrong was writing. Some examples are: Barron and Klein (2016); Godfrey-Smith (2017; 2019); LeDoux (2019); Polger (2017); Polger and Flanagan (2002); Robinson, Maley, and Piccinini (2015); and Rosenthal (2008). Beyond a few comments, I will concentrate here on Armstrong's own view, and the way it interacts with other parts of his philosophy of mind, rather than attempt to connect what he says with this literature.

> mind... Introspection is therefore a logical precondition of teleological mental behaviour. So we have given a 'teleological deduction' of the existence of introspection. (*MTM*, 163)

It is reasonable to assume that Armstrong's account of the function of consciousness, whatever it is, is contained in these passages. So, what is that account, exactly?

I think it is fair to summarize the account as an argument for the conclusion that consciousness has a biological function. The argument may be formulated this way:

P1. Purposive mental activity has a biological function.
P2. Consciousness is necessary for purposive mental activity.
C. Ergo, consciousness has a biological function.

Let me begin by making some comments about the constituent notions here and about the initial plausibility of both the premises and the inference.

3. Commentary

Comment 1. By 'a purpose', as we have seen, Armstrong means an 'information-sensitive mental cause' (*MTM*, 139). To illustrate, suppose that my purpose is to get a drink, or, as we might say more readily these days, I desire to get a drink. This desire is a mental cause, for Armstrong, because it is a state apt to bring about, i.e., cause, certain behaviours: the desire for a drink will usually cause me to get a drink. The desire is information-sensitive, in that its existence and nature will be influenced by what other information I have, that is, by what other beliefs and desires I have. Suppose I learn, and so come to know, that I have a drink, for example, and for some reason hadn't noticed this before. Then my desire to get a drink will, other things being equal, go away or be extinguished. Being extinguished by the knowledge that it is satisfied is one way for a desire to be information-sensitive.

Of course, not all desires are sensitive in this simple way. Suppose I desire to be warm; that is, suppose I desire to be in some persistent state—a 'continuing condition' Armstrong calls it (*MTM*, 140)—rather than for some event to occur. This desire will not go away if I learn that I am warm. Nevertheless the desire is still sensitive to information in a more general sense. If I desire to be warm, and learn that I am, I will normally be disposed to form new desires: the desire to remain warm, for example.[2]

---

[2] Armstrong emphasizes the information-sensitivity of purposes or desires, but presumably this is a feature of mental states in general. Take the belief that I have a drink in my hand. If I learn that I do not have a drink in my hand, this belief will normally go away. In that sense, beliefs are information-sensitive too, since they are usually affected by whatever else I know.

Comment 2. By ‘purposive activity’, Armstrong means a series of actions initiated and sustained by a purpose, in the sense just introduced. As we will see, it is crucial for him that such activity comes in two varieties: (1) purposive mental activity; and (2) purposive physical activity.

Purposive physical activity is a series of physical actions initiated and sustained by a purpose. As an example of purposive physical activity, Armstrong gives a case in which, as he says (*MTM*, 143), ‘I set my face toward the public house’—that is, I decide to go out to the pub, the Nag’s Head Hotel, for example. In such a case, my desire causes a series of physical actions—leaving my front door, walking down Glebe Point Road, turning right on St John’s Road, and so on.

Purposive mental activity is a series of mental actions initiated and sustained by a purpose. As an example of mental purposive activity, he gives a case of mental arithmetic, in which ‘we are doing a calculation “in our head”’ (*MTM*, 327). In such a case, my desire, for example, to arrive at the product of two other numbers, requires me to focus first on certain operations, then remember the outcome of one operation, while performing another, and so on. A different case Armstrong gives is one in which we go through various possible actions in imagination, in order to work out their consequences and so their desirability (see, e.g., *MTM*, 158ff.). I may know there are two routes to the Nag’s Head, for example, and consider in imagination first taking one, then the other. To the extent that the first route is imagined to lead to a negative outcome—e.g., running into a certain moustachioed logician outside the Habit Wine Bar—I may decide on the second.

Comment 3. When he speaks of the ‘biological function’ of consciousness and other mental states, I take Armstrong to be committed to a kind of *adaptationism* about the mind, i.e., he is assuming that, for example, consciousness is a trait whose current presence in a population has a quite specific historical explanation in terms of natural selection. Admittedly, Armstrong is not terribly explicit on this point, and *MTM* predates the contemporary discussion of adaptationism.[3] But this is nevertheless the most plausible interpretation of what he says. After all, whether consciousness is an adaptation is ‘the question asked again and again by so many biologists and psychologists’. And it is adaptations, and not functions of other sorts, about which people tell just-so stories.[4]

Comment 4. Putting comments 1–3 together, P1 makes an empirical claim about the biological past, since it claims that purposive activity played a distinctive role in the evolutionary history of conscious creatures.

What is Armstrong’s evidence for such a claim? The only relevant passage I know of in *MTM* is this:

---

[3] The classic paper here is Gould and Lewontin (1979); see also Orzack and Forber (2017).

[4] This is certainly true of the work referred to in note 1, for example.

> Let us begin by considering what is the biological function of mind. Mental processes, biologically speaking, are those that fall between the stimulus and response. Their biological function is to make the response more sophisticated, and so more efficient. Now, if what goes on in the mind... has a directed or purposive character, this may well aid the agent to make the best response. For instance, in a problem situation, various possible responses may be tried out 'in the imagination' in order to see which response will best fulfil the agent's purposes. As a causal result, the response may be much more efficient.
>
> (*MTM*, 163)

Unfortunately, this passage raises more questions than it answers. Make the response more sophisticated than what? Sophisticated in what way? Why does increased sophistication generate greater efficiency? Did greater efficiency *in fact* play any role in the biological past of human beings? And greater efficiency than what exactly? Without an answer to these questions, it is hard to avoid the conclusion that the passage, and *MTM* more generally, provide no support at all for P1.

If *MTM* offers no support for P1, one might be tempted to dismiss the argument set out above. One might also be tempted to object that the argument cannot possibly capture the account presented in *MTM*. But I think both reactions are mistaken. First, while I have presented Armstrong's account as an argument, I don't mean that he offers an explicit defence of its first premise. On the contrary, I am sure he is not trying to provide an account of the function of mental activity or of the mind in general; rather, he is providing an account of the function of consciousness, *given* the premise that purposive mental activity has a function. Second, while Armstrong does not defend that premise, one might think it fair enough to assume that it is true, especially if we go along with the adaptationism I am assuming Armstrong is working with. After all, adaptationism presumably entails that if we have a thing of sufficient complexity, we may provisionally suppose it to have some biological function, even if we don't know what that function is. Moreover, as we will see later, by purposive mental activity, Armstrong means something similar to what other writers refer to as rationality or reasoning, and it is not at all implausible that rationality or reasoning has a biological function.[5] I propose, therefore, not to reject Armstrong's argument simply because he himself provides no argument for P1.[6]

---

[5] A recent prominent discussion of this issue is Mercier and Sperber (2017).

[6] In this sense, Armstrong's account is an instance of what Robinson, Maley and Piccinini (2015) have recently called the 'strong reading' of an adapationist account. They characterize an account of this sort as follows:

> [o]n a strong reading... consciousness is necessary for certain mental processes. That is, the relevant mental processes cannot take place without consciousness. If so, then it is impossible to have nonconscious functional duplicates of conscious beings. Instead, there must be functions performed by conscious beings that a nonconscious being could not perform. In

Comment 5. By 'consciousness', Armstrong means, as he says, 'the awareness of our own state of mind' (*MTM*, 95). He goes on: '[t]he technical term for such awareness of our own mental state is "introspection" or "introspective awareness"'. It is in this sense that consciousness and introspection come together for Armstrong, and this is why he is happy to move back and forth from one to the other. For him, one is in a conscious mental state if and only if one is introspectively aware of one's being in that state; to put it differently, one is in a conscious mental state if and only if one is conscious in a certain sort of way of being in that state.

The notion of consciousness Armstrong adopts here is what Norman Malcolm in this later debate with Armstrong (see Armstrong and Malcolm 1984) called the 'transitive sense', in which one is conscious *of* something. But Armstrong also uses 'conscious' in an intransitive sense to apply to the state in question. In this sense, a state is (intransitively) conscious if and only if the subject of the state is aware in some way (i.e., conscious in the transitive sense in some way) of being in that state. Since these notions are inter-defined, for most purposes, it makes no difference which one we have in mind.

Comment 6. Armstrong focuses on a notion of consciousness tied to introspection—introspective consciousness, as Lycan says—but he does not deny that there are other notions too; indeed, his later work on this topic prefigures the pluralism about consciousness that is dominant in the literature today (see Armstrong 1980; Block 1997; and for more recent discussions, see Pautz and Stoljar 2019). In that later work, for example, Armstrong distinguishes perceptual consciousness, in which you are conscious of what you perceive, i.e., physical objects and their perceptually available properties, and introspective consciousness, in which you are conscious of, for example, your perceiving something.

Nevertheless, introspective consciousness is, in his view, the central or basic case, the sort of consciousness that explains the rest. This is supposed to follow from reflection on the case of the long-distance truck driver, who figures prominently in Armstrong's later discussions of consciousness (see Armstrong 1980; 1984), but who makes his first oblique appearance in *MTM*:

> This is something that can happen when one is driving very long distances in monotonous conditions. One can 'come to' at some point and realize that one has driven many miles without consciousness of the driving, or, perhaps, of anything else. One has kept the car on the road, changed gears, even, or used the brake, but all in a state of 'automatism'. (*MTM*, 93)

---

order for such claims to stand, a nonconscious organism must be unable to perform the same function; otherwise consciousness would not be needed to obtain the adaptive advantage. (2015, 368)

As I understand it, Armstrong's reasoning concerning this example is as follows: consciousness in the central sense is what is missing before the truck-driver comes to; what is missing before the truck-driver comes to is introspective consciousness; hence introspective consciousness is consciousness in the central sense.

I think—unoriginally, see Lycan and Ryder (2003)—that this reasoning is unpersuasive. While introspective or second-order consciousness *is* missing in the truck-driver case, so too is consciousness in a first-order sense. Armstrong is right that, prior to coming to, the truck-driver fails to be introspectively aware of his own mental states. But the driver also fails to attend to the road or the gears or the brakes, even though presumably he was aware of such things. If you think, as many do, that attention, or attention to a sufficient degree, is one of the factors that turns mere awareness into consciousness in at least one good sense, the truck-driver example does not show the centrality of Armstrong's introspective notion, as opposed to an alternative first-order attention-based one.

Comment 7. While Armstrong overreaches on the centrality of introspective consciousness, I think we should ignore this too in our investigation of the argument presented above. For suppose introspective consciousness, as Armstrong defines it, is merely one of several different varieties of consciousness, rather than the main or basic one; it is still reasonable to focus on the function of consciousness of this variety.

The reason is that this version of consciousness raises in a very straightforward way the very question that Armstrong says is asked again and again, namely, the question of the function of consciousness. What prompts that question is the apparent possibility that various cognitive functions can go on in the absence of consciousness: on the face of it, one can believe, remember, want, perceive, and act, and yet one does not *consciously* believe, remember, want, perceive, or act. This concern is just as urgent in the case of introspective consciousness as in the case of varieties.[7] In sum, what Armstrong is trying to establish is the interesting conclusion that consciousness in his sense does not come apart from some other cognitive functions; it does not detract from the interestingness of that conclusion that there are analogous issues about consciousness in other senses.

Comment 8. 'Necessity', as it occurs in P2, is intended to capture Armstrong's repeated use of modals in the passages above. Take his remark, '[i]f there are to be purposive trains of mental activity, then there *must* equally be some means by which we become apprised of our current mental state' (*MTM*, 327, emphasis added). I take 'must' here to have wide-scope over the conditional; hence what is being said is that, necessarily, if there is purposive mental activity, there is a means by which we become apprised of our current mental state, i.e., there is consciousness. Further, I take 'necessarily' here to be logical or metaphysical necessity, i.e., necessity in the widest sense—it is hard to interpret Armstrong's use of words like 'deduction' in any other way. Hence what is being said is that, in all possible worlds, if there is purposive mental activity, so too there is consciousness.[8]

---

[7] Contemporary work is often focused, not on introspective consciousness, but on phenomenal consciousness (as Block 1997 calls it). Nevertheless, questions about the evolution of consciousness are still prompted by the idea that consciousness in this sense may be removed or altered even though the function may remain.

Comment 9. Putting comments 4–8 together, P2 has nothing to do with the contingent causal history of the world, and *a fortiori* nothing to do with the evolutionary past of conscious creatures. Rather, it is a modal claim: necessarily, if you undergo purposive mental activity, the states in question must be conscious in the introspective sense.

Why does Armstrong believe P2? The basic idea, as I understand it, concerns the distinction between purposive mental activity and purposive physical activity introduced above. In the case of purposive physical activity, we need to keep track of what is going on in the world, and adjust our mental states accordingly. In the case of purposive mental activity, by contrast, we need to keep track of what is going on in the mind. But keeping track of what is going on in the mind is a matter of being introspectively aware of what is going on in the mind, and this in turn requires consciousness. Later we will consider whether the contrast between purposive physical activity and purposive mental activity is really as Armstrong says it is.

Comment 10. We have seen the reasons for supposing that both premises of this argument are true; do they entail the conclusion?

The answer is no. Suppose the heart has a biological function, to pump blood, say; and suppose that, necessarily being located in space is necessary for pumping blood. It does not follow that being located in space has a biological function: being located is a precondition of something that has a function, but does not have the relevant function itself. More generally, A can have a biological function, and B can be necessary to the existence of A, and yet B might fail to have any function. Hence Armstrong's premises do not entail their conclusion.

In view of the fact that the argument is invalid, one might yet again be inclined to reject it. But again I am not going to make this criticism in what follows. For one thing, it may be that, while the premises do not entail the conclusion, they nevertheless support it; for example, perhaps the best explanation of the two premises is that consciousness has a function. Moreover, it may be that the premises can be enriched in such a way that the conclusion does follow; for example, it might be that what is true is not simply that being conscious is necessitated by purposive mental activity but that it explains why purposive mental activity has the function it has.

---

[8] I will consider later whether a weaker notion of necessity would suit Armstrong's purposes.

4. Perceptual and Non-Perceptual Theories of Introspection

So far, I've set out Armstrong's just-so story as an argument for the conclusion that consciousness has a biological function. And I've noted that, while one might question both the first premise of that argument and its passage to the conclusion, this will not be my approach. For the rest of the chapter, therefore, I will focus on Armstrong's second premise—i.e., P2—the premise that consciousness is necessary for purposive mental activity.

In a moment I will consider whether P2 is persuasive on its own terms. But first, I want to point out that Armstrong's commitment to it creates difficulties for what has become the standard way of drawing the distinction between perceptual and non-perceptual theories of introspection.

We have seen that the notion of consciousness for Armstrong is introspective consciousness: a state is conscious in this sense if and only if the subject of the state is aware of being in it in the introspective way. But what is the introspective way? Armstrong himself offers a distinctive answer to this question. For him, the subject needs to be aware of the state in a *perceptual* way. It is for this reason that Armstrong holds an *inner sense* view of introspection.

What is it for a person to be aware of something in a perceptual way? For Armstrong, to perceive something in general is to be disposed to form a certain belief about it. This is quite different from other theories of perception that tend to see perception as a state related to but distinct from belief. On relational views, for example, perception is a relation between a person and a physical object. And on contemporary representational views, even when they trace their origins to *MTM*, perception is often thought of as a state whose functional role is distinct from that of belief.[9]

In light of the fact that the relevant notion of perception is so belief-like for Armstrong, one might suspect that the distinction between his own perceptual view of introspection and a non-perceptual view will not be worth drawing. But that is not true to Armstrong's rhetoric, which continually emphasizes the notion of inner sense. Nor is it true to the literature surrounding Armstrong's book (and indeed other literature on introspection), which often classifies his view as an inner-sense view and draws a distinction between it and other views.

How, then, is the distinction to be drawn? The main suggestion here is due to Sydney Shoemaker (1996, 226–42; 2012, 241–3), who, as I understand him, draws the distinction in terms of the following thesis, which he calls *the impossibility of self-blindness*.

---

[9] See Siegel (2016) and the references therein for some discussion on these distinctions between theories of perception.

(1) Necessarily, if you are in a mental state *M*, and various background conditions obtain, and you are rational, you will believe you are in *M*.

To illustrate the idea behind (1), take the mental state of being in extreme pain. If (1) is true of extreme pain, it is a necessary truth—using again the widest sense of necessity—that if you are in that state, and you are rational and various background conditions obtain, you will believe that you are in that state.

What does Shoemaker mean by 'rationality' and what background conditions does he have in mind? By 'rationality', I take him to be talking about a property of an agent whose combinations of mental states at a time, and whose sequences of mental states over time, conform to the requirements of rationality, where these requirements are understood in a relatively well-known way: other things being equal, do not have overtly contradictory beliefs; other things being equal, believe the obvious consequences of your beliefs; other things being equal, desire to *F* if you believe that *F* is how to do *G*, and you desire to *G*; and so on (see, e.g., Shoemaker 1996, 232; see also Stoljar 2018). As regards background conditions, Shoemaker, I think, has in mind the following: (a) being psychologically capable of forming the belief that you are in pain, that is, understanding the proposition that you are in pain, and being able to believe it; (b) being such that it matters to you that you are in pain—sometimes you might be in pain and be psychologically capable of forming the belief, but it just doesn't matter to you that you are, for example, because you are too busy doing or thinking about other things; (c) having no counter-evidence (provided, for example, by testimony from a friend or a doctor). Putting all this together, it is plausible (whether or not it is ultimately correct) that (1) is true of extreme pain: necessarily, given these conditions, if you are in extreme pain, you will believe that you are, so long as you are rational.

While (1) is plausible in the case of extreme pain, it is not plausible in case of logically analogous conditions. Consider, for example, extreme debt. Even if counterparts of Shoemaker's background conditions hold, and you are rational, it does not follow that you will believe you are in extreme debt if you are. You might be on your way to meet your accountant, for example, and have not yet heard the bad news. If so, (1) is not true of extreme debt. (Remember that (1) asserts that something is necessarily true, and 'necessity' here is to be understood, again, in the widest sense.)

Extreme debt is not a psychological condition, but there are psychological states that behave more like it than like extreme pain. Consider the state of knowing when and how to contract 'want' and 'to' to 'wanna' in English. According to linguists, speakers of English know when and how to do this. They know it is appropriate, for example, in the sentence, 'Who does Mary want to go to the game with?' but they will resist it in the (superficially similar but structurally different) sentence, 'Who does Mary want to go to the game with her?'. (The sentence, 'Who does Mary wanna go to the game with?' is a normal part of speech, but 'Who does Mary wanna go to the game with her?' seems ungrammatical.)[10] Presumably there is a rule that captures when and how speakers do this, a rule they both know and abide by. But it is not plausible that speakers of English know that they know this rule, nor is it plausible that they are able to attend to the rule or to formulate it in words. Likewise, it is not plausible that they can teach it to each other in any ordinary sense of 'teach', since that would require knowing, attending to, and formulating the rule. (Third parties, such as linguists, may know the rule, of course, or at least may hope one day to figure it out; a classic paper is Lasnik and Saito (1984); see also Boeckx 2000.) But suppose now you are an English speaker and thereby abide by the rule in question—call it *R*, whatever it is. It may be that you know *R*, that variations of Shoemaker's background conditions may obtain, and that you are rational, and yet you may nevertheless not believe that you know *R*—you may lack certain crucial pieces of evidence that you know to be relevant, for example. If so, (1) is not true of knowing rule *R*.

How does the thesis of the impossibility of self-blindness—that is, (1)—serve to distinguish anti-perceptual theories from perceptual theories of introspection? Shoemaker's suggestion, I think, is that anti-perceptual theories entail that (1) is true with respect to a range of mental states, whereas perceptual theories entail it is false for all. In other words, perceptual theories want to say that there is a sense in which every psychological state behaves like extreme debt or knowing the rule of 'wanna'-contraction in English: we can come to know about them alright, but, from an epistemological point of view, the way in which we do so is no different from the way in which we know many other things.

Shoemaker motivates this suggestion in part by contrasting (1) with various stronger theses that Armstrong himself criticizes in Chapter 6 of *MTM*. One such thesis is:

(2) Necessarily, if you are in a mental state *M*, you will believe you are in *M*.

If (2) is true of extreme pain, then, if you are in extreme pain, you will believe that you are. Armstrong argues against this by pointing out that the realizer of extreme pain is modally distinct from the realizer of the belief that you are in extreme pain. Like many others, I find this argument unpersuasive. Suppose S_1 is physical state that is the realizer of extreme pain, and S_2 is a physical state that is the realizer of believing that you are in extreme pain. It may be that S_1 and S_2 are modally distinct in the sense that it is possible that you are in S_1 and not in S_2. But it does not follow

---

[10] The main structural difference between these sentences might be brought out as follows. 'Who does Mary want to go to the game with?' has the form, 'Who is the *x* such that Mary wants to go to the game with *x*?' while, 'Who does Mary want to go to the game with her?' has the form, 'Who is the *x* such that Mary wants *x* to go to the game with her?'. In the first case, Mary is the implicit subject of the infinitive 'to go to the game'; in the second, she isn't.

(even if it is true) that it is possible to be in extreme pain and not believe that you are. After all, it may be that, in the situation in which you are in S_1 and not in S_2, S_1 is no longer the realizer of pain! Conclusion: that the realizers of various states are modally distinct does not entail that the realized states are modally distinct.[11]

Even if Armstrong's specific argument against (2) is unpersuasive, he is nevertheless right to reject it, even for the case of extreme pain. If (2) is true, any agent who is in extreme pain will believe that they are. But this is implausible: the cognitive requirements of being in pain and of believing that one is in pain are quite different, which strongly suggests that the first may be met without the second being met. If so, (2) is not true even in the case of extreme pain. But, of course, even if (2) is false, it might be that a weaker thesis, such as (1), is true in that case. That is, in fact, Shoemaker's suggestion.

In summary, it is plausible that Shoemaker has drawn a clear distinction between perceptual theories and non-perceptual theories of introspection. But now the problem that I want to identify for Armstrong's account of the biological function of consciousness emerges. For look again at Armstrong's second premise, P2. That premise says that if you engage in purposive mental activity, you will need to be conscious of the constituent states of that activity, at least in certain cases. The problem is that, on the face of it, Armstrong's purposive mental activity includes Shoemaker's rationality. When Armstrong talks, for example, of adjusting 'mental behaviour to mental circumstances', it is hard to understand what he says, unless he is talking about rational chains of mental states, rational combinations of mental states unfolding over time. But this is precisely what Shoemaker has in mind when he talks of 'rationality'. However, if it is the case that such rational adjustment of mental behaviour to mental circumstances requires consciousness, and if consciousness is just the belief that you are in the relevant state, then it would seem that, according to Armstrong, (1) is true of the purposes involved in mental activity: they are such that, if you are rational, and if various background conditions apply, you must believe that you have them if you do. But if (1) is true of purposes, then Armstrong stands revealed as holding a non-perceptual theory of introspection. Despite what he says, Armstrong is not an inner-sense theorist, at least if we operate with Shoemaker's test for being an inner-sense theorist.

There are various ways that one might respond to this result. First, one might suggest that the distinction between perceptual and non-perceptual models is without foundation and that, properly understood, there is no opposition here: Armstrong wants to reject (2) or something like it, Shoemaker wants to reject (1) or something like it. However, while there is certainly some truth to this, it is hard to see it as an attractive response for Armstrong, who, as I have noted, continually emphasizes the perceptual character of introspection.

---

[11] The classic presentation of this point is the last few paragraphs of Lewis (1999); see also Shoemaker (1996). For some more recent discussion, see Stoljar (2018).

Second, one might suggest that what is really important to Armstrong in this part of his work is his well-known hypothesis of the self-scanner, that is, the idea that, as a matter of fact, there is a causal system instantiated in the brain that reliably permits it to record its own states. However, while Armstrong certainly is committed to this, it is not *all* he is committed to. For by itself the existence of a self-scanner is consistent with a denial of a perceptual theory—at least, if we operate with Shoemaker's account of the perceptual/non-perceptual distinction.[12] But as we have seen, Armstrong repeatedly emphasizes that his own account is a perceptual one.

Third, one might draw the distinction between perceptual and non-perceptual theories of introspection in a way different from Shoemaker. For example, one might argue that perceptual theories portray introspection as like perception, not on a representational theory of perception, but instead on a relational theory. However, even if this can be done, it is again unclear that it will help Armstrong. The reason this time is that it is probable that Armstrong will turn up on the wrong side of the divide. So once again we seem to confront the uncomfortable truth that Armstrong is not the perceptual theorist he claims to be.

Finally, one might try to respond to the point by downplaying it, that is, by suggesting that purposive mental activity is an unusual phenomenon. I suspect this would be Armstrong's own response to the issue, but the problem with it is that purposive mental activity is not unusual. In fact, any sort of purposive activity at all will include purposive mental activity of some sort or another. However, this point will emerge more naturally when we turn to the second criticism of Armstrong's position that I want to develop.

5. Two Notions of Information Sensitivity

We have seen that Armstrong's argument about the function of consciousness has a surprising side effect: its second premise entails that he does not hold an inner-sense theory of introspection, at least on Shoemaker's understanding of what it is to hold such a theory. But, of course, this point is merely *ad hominem*; it does not by itself say much about whether the premise is true. So that is the question I want to consider now.

---

[12] Indeed, Shoemaker makes this point himself in this discussion of Armstrong's broad perceptual model, particularly when he separates what he calls the 'causal condition'—roughly, the idea of a self-scanner—from the 'independence condition'—roughly, the idea that self-blindness is possible; see Shoemaker (1996, 224).

As we saw earlier, the consideration Armstrong offers for P2 is that purposes are information-sensitive mental causes, and that information sensitivity requires consciousness. However, the problem with this line of thought is that there are two ways to interpret the notion of information-sensitivity: a first-order way and a second-order way. To illustrate, consider again the desire to get a drink. On the first-order interpretation, this state is information-sensitive just in case if you learn you have a drink, the desire to get a drink will go away, other things being equal—we may put this by saying the state has first-order information sensitivity. On the second-order interpretation, by contrast, the state is information-sensitive just in case if you learn that you have a drink, and *if you are introspectively aware of learning this*, the desire will go away, other things being equal. Or consider a non-purposive state that is presumably also information-sensitive, such as the belief that p. This state is first-order information-sensitive (as we may say) just in case, if you learn that p is false, the belief that p will go away, other things being equal. By contrast, the state is second-order information sensitive just in case, if you learn that p is false, *and if you are introspectively aware of learning this*, the belief that p will go away, other things being equal.

Once we have this distinction before us, however, the problem for P2 is straightforward. If we understand information-sensitivity in the first-order way, P2 is false: it is not true that introspective consciousness is necessary for purposive mental activity, for one can engage in such activity without being aware of the relevant states. If we understand it in the second-order way, however, then P2 is unjustified, since none of Armstrong's examples motivate the idea that mental states are information-sensitive in this second-order way.

In the case of mental arithmetic, for example, it is true that one's next step depends in an information-sensitive way on the step before, but there does not need to be introspective consciousness at any point; on the contrary, in such a case one seems firmly focused on mathematical rather than psychological issues. The same thing is true in the case in which I imagine two routes to the Nag's Head. I might consider taking the first route, which leads past the Habit Wine Bar, and resolve after deliberation to take the longer route via Ross St. No introspection need occur here at all. It is true that there needs to be information-sensitivity in the sense that the mental states will fall into a rational sequence, but nothing beyond that is required.

Indeed, the point that nothing in these examples demonstrates the need for second-order information-sensitivity can be strengthened if we bring to the surface what seems to be a false presupposition in Armstrong's entire discussion of the distinction between mental and physical purposive activity. As we have noticed, Armstrong does not think that consciousness is necessary for physical purposive activity; it is only necessary for mental purposive activity. But this betrays a misunderstanding of this distinction. There are trains of purposive activity concerning many different things: mathematics, pubs, logicians, moustaches, and so on. Some issue in physical activity, some don't. But almost all of these involve mental activity of some sort, and none of them requires the mental activity or the states that constitute this activity to be conscious. If not, there is no support here for P2.

At certain points, Armstrong himself comes close to seeing the basic problem with his account. In response to his own discussion, for example, he points out that '[o]n occasion, we can solve quite complex problems during sleep, or while our mind seems to be otherwise occupied. If this is possible, where is the need for awareness of our own mental state in purposive mental activity?' (*MTM*, 164). That is a good question, but Armstrong's response is implausible. What he says is that in such cases we are aware, but not aware that we are aware. In other words, his claim is that in such cases we have a lack of third-order awareness but not a lack of second-order awareness. But while it is indeed plausible that in deep sleep we do not have third-order awareness, this does little to undermine the very plausible suggestion that we lack second-order awareness too.[13]

I have argued that P2 as Armstrong formulates it, and therefore his argument in general, are unpersuasive. Is it possible to adjust the premise so that P2 becomes more plausible?

One possibility is to focus on the strength of the necessity claim in P2. Above, I interpreted it widely, as involving all possible worlds. But one might try to interpret it more weakly. On this interpretation, P2 says not that consciousness is metaphysically necessary for purposive mental activity but only that it is empirically necessary for it.

However, the problem with this proposal is that, if Armstrong's account is developed in this way, the speculative ('just-so') element in his position becomes extreme. Earlier I said that Armstrong's account was premised on the empirical assumption the mental activity had a function, noting that this was a reasonable assumption in the circumstances. But if his account is premised on the speculation that consciousness is required for mental activity to perform its function, whatever it is, we lose the sense in which he has an answer to the question of the evolution of consciousness at all.

A different possibility is to adjust the notion of 'purposive mental activity' so that what is intended is purposive *conscious* activity: a series of conscious states initiated and sustained by a purpose. However, the problem with this proposal is that while P2 is now plausible, P1 begs the question. From this point of view,

---

[13] Might it be insisted on Armstrong's behalf that, since second-order awareness is part of the functional role of purposive mental activity, if you don't have awareness of that sort, you don't have the activity in question? I think the problem with this suggestion is two-fold. First, it is hard to see it as attractive to Armstrong, who in general insists on the point that introspective awareness can come apart from first-order states. Second, it is not plausible on its own terms, since it remains the case that there is no justification for building this into the functional role of purposive mental activity in the way that he does. I am indebted here to David Braddon-Mitchell.

Armstrong's P1 amounts to the claim that conscious mental activity has a function, and that is precisely the thing the argument aims to establish.

6. The McGinn-Shoemaker Argument

In a certain sense, the problem for Armstrong's account we have just been looking at was predictable from the beginning. As noted above, the reason that people are interested in the evolution of consciousness is because of the apparent possibility that various mental states, events, and processes can occur without being conscious. What Armstrong is trying to do is rule out that possibility. His idea is that there is a particular sort of mental process that cannot go on without consciousness as a matter of necessity, namely, purposive mental activity. But the problem is that this does not seem to be true; we seem to be able to engage in purposive mental activity *whether or not* we are conscious in Armstrong's sense. For all Armstrong has said, therefore, the question asked again and again remains unanswered.

At this point, various avenues are open to us. One is to leave Armstrong aside and focus for its own sake on what the function of consciousness is. It goes without saying that I am not going to attempt that here. Instead, let me close the discussion by bringing together the two critical points I have made about Armstrong's account: (1) that it undermines his claim to be an inner-sense theorist; and (2) that it conflates two notions of information-sensitivity. The connection between these points comes into sharper relief if we compare Armstrong's discussion with a more contemporary issue in philosophy of introspection.

The issue I have in mind returns us to Shoemaker's thesis of the impossibility of self-blindness, namely (1): necessarily, if you are in a mental state *M*, and various background conditions obtain, and you are rational, you will believe you are in *M*. As I noted, Shoemaker argues for this thesis in various ways—partly on the basis of examples, and partly by comparing it with other stronger principles that are implausible. But he also presents an argument for the thesis that is striking in its similarity to Armstrong's discussion.

The main idea of this argument, which is based on a suggestion due to Colin McGinn, is that in order to abide by principles of rationality, it must be that one knows what belief states one is in (see McGinn 1982, 20; Shoemaker 1996; 2012). Suppose that for purposes of illustration we focus on what might be called a 'no-contradiction rule', namely, it is rationally required that if one believes p, one will not also believe not-p, other things being equal. Then we may present the argument as follows:

(3) If you are rational you abide by rules such as the no-contradiction rule.

(4) If you abide by rules such as the no-contradiction rule, you must know (and so believe) what belief states you are in.

From these premises it is reasonable to infer that (3) is true of any state that falls within the purview of the no-contradiction rule, at least, if the premises themselves are necessary.

But the weakness in this argument is that (4) is unconvincing. If one abides by the no-contradiction rule, it is true that what belief states one is in need to be changed over time, revised, regulated, and so on. However, it seems perfectly possible that such revision goes on, as Richard Moran puts it in an important critique of this line of thought, at a sub-personal level: '[t]he purposes of belief-regulation do not require that the *person* get involved here at all' (Moran 2001, 110).

Now, in response to Moran, Shoemaker has suggested that, while this might be true in some cases, it is not true in all:

> But I think that in an important class of cases the revision and updating do require that there be second-order beliefs about what the contents of the belief system are. These are cases in which the revision of the belief system requires an investigation on the part of the subject, one that involves conducting experiments, collecting data relevant to certain issues, or initiating reasoning aimed at answering certain questions. Such an investigation will be an intentional activity on the part of the subject, and one motivated in part by beliefs about the current contents of the belief system. These will include the belief that there are certain apparent inconsistencies or incoherencies in the system, the belief that there are gaps in how the system represents the world, and the realization that the system represents the existence of certain states of affairs for which it provides no explanation. (Shoemaker 2012, 244)

However, a version of Moran's objection applies here too. It is true that if one embarks on the sort of investigation Shoemaker describes, it is irrational to believe the relevant proposition. But that implies only that a rational agent who undergoes this sort of investigation must somehow or other fail to believe that proposition—either they never did believe it at all, or else they will suspend belief in it for the course of the investigation. But this failure to believe is something that, on the face of it, could be brought about in the way Moran suggests, i.e., sub-personally. If so, there is no reason to accept Shoemaker's claim that this sort of investigation is motivated by beliefs about the current content of the belief system.

No doubt there is more to say about the dispute between Shoemaker and Moran on that point. I will not discuss it further here, except to note that my sympathies are with Moran. My main point is to notice that Shoemaker's argument here is strikingly similar to Armstrong's and yet is used in the service of a non-perceptual theory rather than a perceptual one; further evidence, then, that the lines between these two are starting to blur. Likewise, the problem with Shoemaker's argument is extremely similar to the problem we isolated for Armstrong's. Moran himself says that the 'person need not get into the picture'. That's one way of putting it. Another is that abiding by the no-contradiction rule requires first-order but not second-order informational sensitivity. If so, both Shoemaker's and Armstrong's positions are mistaken and for the same reason.

7. Conclusion

Our overall discussion may now be summarized as follows. (1) Armstrong's just-so story about the function of consciousness is a revealing and under-noticed part of *MTM*. (2) Its second premise, P2, undermines the distinction between perceptual and non-perceptual theories of introspection. (3) P2 is false since it neglects the possibility of a first-order approach to information sensitivity. Finally, (4) further support for the previous two points can be found by noticing that Armstrong's discussion is strikingly similar to Shoemaker's discussion.

References

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1973) *Belief, Truth and Knowledge*, London: Routledge.

Armstrong, D. M. (1980) *The Nature of Mind and Other Essays*, Ithaca, NY: Cornell University Press.

Armstrong, D. M. and Norman Malcolm (1984) *Consciousness and Causality: A Debate on the Nature of Mind*, Oxford: Blackwell.

Barron, Andrew B. and Colin Klein (2016) 'What Insects Can Tell Us about the Origins of Consciousness', *PNAS*, 113(18): 4900–8.

Block, Ned (1997) 'On a Confusion About a Function of Consciousness', in *The Nature of Consciousness*, edited by Ned Block, Owen J. Flanagan, and G. Guzeldere, Cambridge, MA: MIT Press.

Boeckx, Cedric (2000) 'A Note on Contraction', *Linguistic Inquiry*, 31: 357–66.

Godfrey-Smith, Peter (2017) *Other Minds: The Octopus, the Sea and the Deep Origins of Consciousness*, London: HarperCollins.

Godfrey-Smith, Peter (2019) 'Evolving Across the Explanatory Gap', *Philosophy, Theory, and Practice in Biology*, 11: 1. DOI: https://doi-org.ezproxy.library.sydney.edu.au/10.3998/ptpbio.16039257.0011.001

Gould, Stephen Jay and Richard C. Lewontin (1979) 'The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme', *Proceedings of the Royal Society of London, Series B*, 205: 581–98.

Lasnik, Howard and Mamoru Saito (1984) 'On the Nature of Proper Government', *Linguistic Inquiry*, 15: 235–89.

LeDoux, Joseph (2019) *The Deep History of Ourselves*, New York: Penguin.

Lewis, David (1999) 'Psychophysical and Theoretical Identifications', in *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press, 248–61. First published in *Australasian Journal of Philosophy*, 50, 1972: 249–58.

Lycan, William G. (1996) *Consciousness and Experience*, Cambridge, MA: MIT Press.

Lycan, William G. and Zena Ryder (2003) 'The Loneliness of the Long-distance Truck Driver', *Analysis*, 63(2): 132–6.

McGinn, Colin (1982) *The Character of Mind*, Oxford: Oxford University Press.

Mercier, Hugo and Dan Sperber (2017) *The Enigma of Reason: A New Theory of Human Understanding*, Cambridge, MA: Harvard University Press.

Moran, Richard (2001) *Authority and Estrangement: An Essay on Self-Knowledge*, Cambridge, MA: Harvard University Press.

Orzack, Steven H. and Patrick Forber (2017) 'Adaptationism', in *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/spr2017/entries/adaptationism/

Pautz, Adam and Daniel Stoljar (2019) *Blockheads! Essays on Ned Block's Philosophy of Mind and Consciousness*, Cambridge, MA: MIT Press.

Polger, Thomas W. (2017) 'Rethinking the Evolution of Consciousness', in *The Blackwell Companion to Consciousness*, 2nd edition, edited by Susan Schneider and Max Velmans, Oxford: Blackwell, 77–92.

Polger, Thomas W. and Owen J. Flanagan (2002) 'Consciousness, Adaptation and Epiphenomenalism', in *Consciousness Evolving*, edited by James H. Fetzer, Amsterdam: John Benjamins, 21–42.

Robinson, Zach, Corey J. Maley, and Gualtiero Piccinni (2015) 'Is Consciousness a Spandrel?', *Journal of the American Philosophical Association*, 1: 365–83.

Rosenthal, David (2008) 'Consciousness and its Function', *Neuropsychologia*, 46: 829–40.

Shoemaker, Sydney (1996) *The First Person Perspective and Other Essays*, Cambridge: Cambridge University Press.

Shoemaker, Sydney (2012) 'Self Intimation and Second Order Belief', in *Introspection and Consciousness*, edited by Declan Smithies and Daniel Stoljar, New York: Oxford University Press, 239–57. First published in *Erkenntnis*, 71, 2009: 35–51.

Siegel, Susanna (2016) 'The Contents of Perception', in *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2016/entries/perception-contents/

Stoljar, Daniel (2018) 'Introspection and Necessity', *Nous*, 52: 389–410.

11

Armstrong on Conation

William G. Lycan

It is a great honour to participate in this celebration. And: I'm so old that I remember the book's original appearing. In fact, I remember it very well, because at the time, I was a graduate student, and the Chicago philosophy faculty had been keenly aware that *A Materialist Theory of the Mind* (hereafter *MTM*) was on its way; *two* separate seminars were immediately given on it, I think by Leonard Linsky and Vere Chappell, and I was enrolled in each of the two. For one of them, I wrote a paper criticizing Armstrong's general causal-analysis strategy on dialectical grounds. The prof pointed out to me that it was a bad, terrible paper. So I just threw it away; good thing for everyone concerned.

Armstrong's signature way of defending the token identity of mental states with brain states has been discussed at great length by many. Less attention has been given to his many respective causal analyses of particular mental state and event types, that occupy the 223(!) pages of Part Two. (I remember a review of the book, Thomas Nagel's, which observed: 'I am inclined to think that the attempt at comprehensiveness is a mistake. (In the section on emotions, for example, embarrassment gets six lines, disgust five, and despair three.)' (Nagel 1970, 396).

Actually, Armstrong had said that 'the discussion in this part will be confined to those sorts of mental states that anti-Materialists . . . have thought to raise special difficulties for Materialism' (*MTM*, 129). My main task in this chapter is to examine Armstrong's causal theory of desire, though it was not highlighted in the book. I am being slightly selfish, because in the past few years I have been investigating desire considered as a mental state,[1] but it's also true that Armstrong's views on conation have been severely underdiscussed.

There is no mistaking the influence of Ryle. Part Two is deliberately entitled 'The Concept of Mind'. And in Section VI of Chapter 6, Armstrong makes the neat point that we can keep Ryle's view that a mental state is a disposition to behave, *if* we reject Ryle's shallow counterfactual understanding of dispositions and substitute the 'Realist' understanding, according to which a disposition is to be empirically identified with its categorical basis. We then just accept the plausible claim that a disposition's categorical basis is what causes the

---

[1] See e.g., Lycan (2012); incidentally, presented to the Department of Philosophy at Sydney University in July 2009.

William G. Lycan, *Armstrong on Conation* 'Ought' In: *Armstrong's Materialist Theory of Mind.* Edited by: Peter R. Anstey and David Braddon-Mitchell, Oxford University Press. © William G. Lycan 2022. DOI: 10.1093/oso/9780192843722.003.0012

disposition's manifestations in the relevant circumstances, and voilà, Rylean behaviourism collapses into Armstrong's common-sense causal theory.

1. Purposes

Part Two begins, *not* actually with any particular mental state, but, surprisingly to us today, with the concept of *purposive activity*. To the twenty-first-century philosopher of mind, that needs explaining. Nowadays—a mindset that hardened in the 1970s and 1980s—we think in terms of 'belief-desire psychology'. Beliefs and desires are roughly equal partners in producing action, normally by resulting in intentions which produce the actions more directly, and all in terms of propositional-attitude content.[2] But it was not so in 1949, and here too Armstrong follows Ryle's lead.

(For that matter, why does Armstrong *begin* his catalogue with conation? I never asked, but I suspect the reason was that there is a closer conceptual connection between desire or intention and behaviour than there is between belief or perception and behaviour. So to begin with conation helps to soften us up.[3])

It is a little hard to identify the first mental state-type addressed by Armstrong as such. He speaks of 'the Will' as referring collectively to 'every sort of mental process that is of the conative sort' (*MTM*, 131), and he uses the word 'purpose' to refer generically to conative mental causes of 'behaviour proper' (*MTM*, 84, 132). For him, a paradigmatic sort of purpose is an intention, but Armstrong does not say that every piece of behaviour proper is caused by an intention; in fact, later (*MTM*, 143) he explicitly denies that. Indeed, he grants that not even purpose in the ordinary sense is required: a person may 'raise . . . his arm . . . simply because it comes into his head to do so' (*MTM*, 131–2), and that action 'may involve no purpose, not even the purpose of raising his arm' (*MTM*, 133); it may be '*quite* idle' (*MTM*, 133, italics original). But Armstrong continues to use the word 'purpose' in an artificially broader sense, as a placeholder for whatever sort of conative cause is at work in a given case.

Not just any mental cause counts as a purpose, even in that broad sense; nervousness causes thumb twiddling, sudden joy may cause a delighted gasp, and fear may cause a start. Building on his analogy with the homing rocket (*MTM*, 138), Armstrong develops the idea that what marks a mental state as a purpose is its *information-sensitivity* (*MTM*, 139): The mental cause is identified in terms of a goal or objective (which explains both the intensionality and the

---

[2] That picture came from Davidson (1963) and later, independently, from Dennett (1971). Peter Godfrey-Smith reminded us that it had been explicit in Ramsey, but although we know Armstrong read Ramsey, it did not take.

[3] The same had been true of Ryle; following his general Chapter II, on knowing how vs. knowing that, his first more specific chapter was Chapter III: 'The Will'.

Brentano-intentionality of its ascription, *MTM*, 140, 144). And the cause is affected by *perceptions* 'or some other form of awareness' (*MTM*, 140) *en route* to the goal; in response to new information, it makes changes of detail that correct the agent's course and keep her/him on track. Armstrong says the mental cause itself is 'modif[ied] . . . in a way suitable (or believed by the agent to be suitable) for the achievement of the "end" of the activity' (*MTM*, 139), from which I infer that he thinks of the cause itself as complex, perhaps having sub-causes arranged in a planning structure of some sort.

Except 'where what is sought is . . . a continuing condition' (*MTM*, 140), the purpose will cease to exist when the end or objective is recognized to have been attained. But it often happens that a purpose will cease for some other reason: 'because circumstances are or seem to be such that it cannot be fulfilled', or 'because a more attractive or compelling line of conduct suggests itself' (*MTM*, 142) (or, I should think, for no psychological reason at all). To accommodate these cases, Armstrong goes counterfactual and notes that when we consider 'differences of circumstances of the sort that might reasonably be expected', we see that there is a 'cluster' point:

> Despite the [hypothetical] changes in circumstances, the line of my life would still have passed through the one point, and perception that I was at that point would have extinguished the mental cause. But now consider those changes in my environment, or in me, which would have led, either voluntarily or compulsorily, to the abandonment of my purpose. The points in time on these hypothetical lines of life where the mental cause ceased to drive would show no such tendency to cluster around the same happening. (*MTM*, 143)[4]

This is what specifies the goal or objective for sure, though I am not yet certain how to fold it more precisely into the main causal analysis.

Now, what about the behaviour that a purpose is apt for causing? (Armstrong's official general formula for the causal analysis of a mental state is: '*a state of the person apt for bringing about a certain sort of behaviour*' (*MTM*, 82, italics original), with the addition that for some mental states we must add '*apt for being brought about by a certain sort of stimulus*'.) An end, goal, or objective is not usually, or anyway not always, a piece of behaviour on the agent's part. Armstrong is oddly silent on this point. But he touches on it in his later paper, 'The Causal Theory of the Mind',[5] actually mentioning desire: '[f]or instance, the desire for food is a state of a person or animal that characteristically brings about food-seeking and food-consuming behaviour by that person or animal' (Armstrong

---

[4] In this, he follows R. B. Braithwaite's (1953) *Scientific Explanation*: 'a tendency to arrive at exactly the same point under diverse possible circumstances' (*MTM*, 143).

[5] First published in 1977 and reprinted in Armstrong (1980).

1980, 21–2). We're supposed to know what sorts of behaviours those are; fair enough, or at least philosophical Behaviourists had done little better in giving independent characterizations of the behaviour to which they appeal, and *they* were trying not to rely on a prior conception of mind.[6] So the answer to our question is: The behaviour that a purpose is constitutively apt for causing is the behaviour that in the context would count as seeking the objective or trying to realize it.

2. Pro-Attitudes and Desires Proper

Armstrong does not offer a causal analysis of intending or of desire *per se*. He makes it clear that each of those states is a subspecies of 'purposes' in his broad sense. A purpose is an intention or intending when 'the objective is something the agent thinks to be within his power . . . at the time of having the purpose' (*MTM*, 149). A desire in the most common sense is a purpose that may not be 'alone in the field' (*MTM*, 153) but must face competition from other desires that may be stronger. (It will still be true of a given desire that, had it been alone in the field, it would have initiated the relevant line of action.) But more on this shortly.

Armstrong provides no summary statement of his causal analysis of purpose, but leaves us to pull it together. He has never spoken of a typical cause for a purpose, so we suppose that the concept will be explicated in terms of just its normal effects, plus the operative conditions and qualifications. It would read something like this:

> S's purpose to achieve A $=_{\text{def}}$ A state of S apt for the production of information-sensitively A-seeking behaviour and with A as a counterfactual cluster point.

S's *intention* to do A would then be, a state of S apt for the production of information-sensitively A-seeking behaviour and with A as a counterfactual cluster point, where S thinks achieving A to be within S's power.

---

[6] Armstrong (1980, 22):

> in the case of mental states, their effects are all those complexities of behaviour that mark off men and higher animals from the rest of the objects in the world. Furthermore, differences in such behaviour are elaborately correlated with differences in the mental causes operating. So it is only to be expected that the causal patterns invoked by the mental concepts should be extremely complex and sophisticated.

In Chapter 5 in this volume, Peter Godfrey-Smith points out that Armstrong makes no concession to the objection famously directed against Rylean behaviourism by Geach and by Chisholm, that a desire issues in a particular piece of behaviour only as mediated by the subject's existing beliefs and other mental states. Back in the 1970s, I had tacitly assimilated Armstrong to Lewis, and supposed that his view with the Geach-Chisholm point incorporated could be expressed in terms of Ramsey sentences, but I didn't see exactly how that would go; I non-urgently brought up the matter in conversation with Armstrong sometime in 1978, but the conversation was inconclusive.

It is a little trickier to extrapolate the analysis to desire, because of an ambiguity in the term—an ambiguity presciently recognized by Armstrong. Nowadays everyone who works in this area agrees that 'desire' has at least two readings. In one sense, it's entirely commonplace that we do things we do not want to do. I often do things I have no desire *at all* to do; I do them against my will, but for what I consider good and overriding reasons. I'd be very surprised if the same isn't true of each of you; a person who does only things s/he wants to do at the time is at least a borderline sociopath.[7] Fred Schueler (1995) calls desires in this sense 'desires proper'; I believe those are what are typically meant by the term.

But there is also a sense in which, if I did do A voluntarily, I must have ultimately wanted to do A; I did, after all, finally choose to do A. Schueler calls desires in this sense 'pro-attitudes'.[8] That is a more general, the most general, conative category; of course, it includes desires proper. Armstrong had recognized that use of the term:

> What we must reject is the idea that, over and above desires, there are further mental states or processes which are responsible for the translation of desires into action. Desires are essentially action-producing although they may be inhibited or prevented from producing action by other desires. In old-fashioned language, the Will is not a separate faculty from Desire. *Our purpose is simply the desire that is dominating our conduct.* (*MTM*, 152, italics mine)

It seems he means that a purpose is always a broad-sense desire, = a pro-attitude. (Just possibly he meant only that when a purpose is a desire at all it is a pro-attitude, but that weaker reading is discouraged by 'the Will is not a separate faculty from Desire'.[9])

---

[7] In my life I have known exactly one person who actually seemed to have this property, a departmental colleague from over forty years ago (*before* my first visit to Sydney). Fortunately for the department and for society in general, this person was also a naturally benign and sympathetic individual and nearly always wanted to help others, to make relationships work, and to contribute to common enterprises. There were spectacular exceptions ('You wouldn't want me to be unhappy, would you?').

[8] Schueler wielded his distinction against Dennett's (1971) simple version of belief-desire psychology, catching Dennett in a dilemma. I think his argument succeeds; Dennett has never replied to it. It also generalizes to other simple versions. The problem is roughly this: if by 'desire' the theorist means desire-proper, then desires are only one kind of pro-attitude; they do not predict action with any reliability and they only sometimes explain it. But if 'desire' means just any pro-attitude, desires cannot in principle predict behaviour and can only vacuously explain it.

[9] The now standard belief-desire picture does maintain that over and above desires, there is (normally or at least often) a further mental state that is responsible for the translation of desires into action, for that is the role of intentions, which are the proximal causes of action. Desires are thought to cause action, when they do, only by way of causing intentions (which of course is not to say that an intention always does cause the action intended, and an action can be 'unintentional' under one or more of its descriptions). In any case, for a more nuanced and detailed view of intention, see Bratman (1987).

He goes on to say: 'The word "desire" may be said to be a concertina-word: it moves between narrower and wider senses . . . We can say: purposes are desires that we act from; or we can say: desires we do not act from are inhibited purposes' (*MTM*, 153–4). That is not yet the distinction that I have mentioned. Both Schueler's and my concern were with, not desires we do not act from, but actions we do contrary to all desires(-proper). Armstrong had addressed that a page or two earlier, and now we come to some disagreement.

3. Acting Contrary to Our Desires

> [W]e do talk of acting in a way contrary to our desires. What do we mean by this? We may merely mean acting contrary to our immoral impulses. But normally, I think, we are referring to a semi-formal distinction in the nature of our desires: between desires that are not prudential and those that are, between desires that take no account of what the results of action will be in the long run and those that do take account, between the warm and the cool passions as Hume put it. When I act contrary to my desires (which I may do for moral, immoral or amoral reasons), some or all of my short-term or impulsive desires are inhibited by desires that do not have this character.
>
> (*MTM*, 152–3)

Thus, Armstrong does not grant my claim that we can act contrary to all desires-proper. *His* weaker sense of the term refers just to 'warm', short-term or impulsive desires, and his concession is only that we sometimes act contrary to all of those in particular.

For some cases, his account is still accurate: I do not want to eat my undressed broccoli and kale salad, but I eat it because I need the good nutrition it furnishes. My desire not to eat it is inhibited by my longer-term desire for health; fair enough. But suppose on another occasion I do not want to eat the broccoli and kale salad, and I care not at all for nutrition, but I eat it because not to do so would be impolite to my host. Armstrong could suggest that I want to be polite, but perhaps I don't much want to be polite either. Do I have a further, more long-term desire-*proper* in aid of which I am doing the polite thing? Maybe I do, but to insist that I must have one would beg the question. Of course, I must have some motivation or other, but the present issue is whether that motivation must in turn take the form of a desire-proper.

The obvious example of a motivating state that may be contrary to all desires-proper is a normative belief, paradigmatically a moral belief. Perhaps I am determined to be polite, not because I want to but because I believe I am obligated, in a moral or a nonmoral way, to do so. If the perceived obligation is nonmoral, Armstrong might still find some general desire-proper that is motivating me. But suppose I am motivated by a moral belief, and solely so. It would be otiose, because empty, to insist that I must have a desire-proper to be moral, even though of course I do desire in the broad, pro-attitude sense to be moral.

Incidentally, there is no case for Hume's view that beliefs alone cannot produce action, because, as just noted, one common pro-attitude that motivates is the belief that one ought (morally or otherwise) to do such-and-such. Since it itself motivates, or so I have argued elsewhere,[10] that belief need not be accompanied by any further pro-attitude such as a desire-proper.[11]

In any case, nothing much hangs on this. If 'purpose' is a placeholder category for whatever more specific types of motivating state there are, and there are several different types of motivating state, whatever they are called, the respectively appropriate differentia can be added to Armstrong's generic causal analysis of purpose. (Though if beliefs can motivate depending on their contents, that would require a revision in Armstrong's causal analysis of belief.)

We still do not know quite enough to state a causal analysis of desires-proper. (That's not an objection, since the category is not one recognized by Armstrong.) Obviously, not all desires-proper produce their objective-seeking actions, though each is apt for doing so. He says we may think of desires not acted on as 'inhibited purposes', but if 'purpose' there has its broad sense, that does not help. We would need to look at other kinds of pro-attitude and contrast their respective causal profiles.

4. Desire Strength

Of course, desires-proper can compete with each other. Like most people, Armstrong speaks of the 'strength' of a desire, but he points out that the term is equivocal, as between strong in the sense of the desire's power to outcompete other motivations in steering the agent toward its constitutive end, and strong to the extent that 'it creates, or tends to create, a condition of bodily agitation' (*MTM*, 153). Those two sorts of strength can come apart. He might have added a third kind or respect of strength: introspective prominence, which does not necessarily track with either of the other two. And it should be noted that both bodily agitation and introspective prominence can contribute to behaviour-causal power; we may want to distinguish original or direct power from indirectly augmented power that results from one or both of the other kinds of strength.

---

[10] 'Moral Facts and Moral Knowledge', reprinted as Chapter 11 in Lycan (1988).

[11] I don't know whether Armstrong would have agreed; I don't recall his joining that debate in print. In meta-ethics his sympathies were with John Mackie's error theory.

Perhaps there is a separate psychological category we could call 'norm-obedience'. In any case, Schueler points out (1995, 31–4) that pro-attitudes *as such* are not what figure as premises in practical deliberation and reasoning. It is *reasons* that do that. In a given case, one's reasons may (or may not) include desires-proper, feelings such as hunger or pain, attitudes of one kind or another, or normative beliefs. (Schueler's example of the last is his belief that he has a responsibility to attend a meeting at his son's school, though he prefers to think of the reason—what would figure as an actual premise in his practical reasoning—as just that he does have that responsibility. What never figures in his reasoning is a desire-proper whose content is: to live up to his responsibilities; notice that that is not a tendentious or even contentious claim.) Practical reasons will be or reference, generically, pro-attitudes; not vice versa.

5. Wants, Wishes, and Hopes

Section IX of Chapter 7 is called 'Wants and Wishes', and Armstrong distinguishes 'wants' from everyday desires. (Certainly some stipulation here.) Wants, he proposes, are desires for 'things that . . . are completely unobtainable, or, if obtainable, are things that we can do nothing to obtain[, such as] . . . desires to live forever, or to know everything there is to know' (*MTM*, 154). He does not distinguish these from wishes. The problem is that they are not causes of action, at least not of seeking their objects as objectives. They 'do not govern our actions, or even press towards governing our actions' (*MTM*, 154).

He offers a nested counterfactual account: To want or wish that P, where P is known to be beyond our power, 'is to be in a mental state such that, if we were to believe the objective of the state could be easily fulfilled by certain courses of action, we would attempt to fulfil it, or, at least, there would be a tendency to attempt to fulfil it' (*MTM*, 155). He gives the example of wishing I could have attended the first night of *Twelfth Night*: '[I]f I had lived in Elizabethan England, but known something of what is known of Shakespeare today, I would have . . .' (*MTM*, 155). That sounds right to me.

Armstrong has hardly anything to say of hopes. This is it: '*Hope* involves a lack of knowledge that something is so, but a belief that it is at least a real possibility, together with a desire that the thing should be so' (*MTM*, 186). Our current empress of hope, Adrienne Martin,[12] calls (roughly) this the 'orthodox' definition, but she argues that *mere* desire is not enough; the subject must adopt a positive normative stance toward the desire. At any rate, she agrees that a hope involves a desire and is not a separate kind of conation.

6. The 'Grain' Problem

Now I must mention an issue that has arisen only in the last decade. A little nest of issues, actually, two of which need not have concerned Armstrong because they are semantical rather than metaphysical, but they are intertwined.

For me, it began with a puzzle I called the 'grain' problem. It was discovered independently by the late Delia Graff Fara; she called it the 'content specification' problem.[13] Suppose you desire to have lunch. I sit you down before a large piece of gross, mouldy cheese excavated from the back of my refrigerator, and I tuck your napkin under your chin. Says you, 'Not *that* lunch! *Not* what I wanted!' Or you desire to be famous. You publish something, but it contains a fallacy so obvious and stupid that the news of your gaffe spreads to the entire English-speaking world, and your photograph appears on posters. 'I didn't mean *that* way/famous for *that*.' The question now is, how does a desire relate its subject to the proposition expressed by the complement sentence of its canonical description?

---

[12] See Martin (2014).

On the face of things, what is desired in the above examples is not precisely what was said to be desired, because is it really much more specific. But, Fara and I agree, the desire ascription is not thereby falsified. It remains true in the context that you wanted to have lunch, or to be famous, despite the need for tacit qualification. The ascription is put categorically and is not overtly conditional. That, we suppose, is because the qualifications are taken for granted. (This does not happen with belief ascriptions, at least not so commonly and in the same way.)

In my previous work I surveyed a number of approaches to this phenomenon. My preferred one is this: The desires in question really *are* conditional desires. Consider the account of conditional desires put forward by Kris McDaniel and Ben Bradley (2008). Their opening examples are: B wants to go out for a beer later, so long as he is not too tired, or he wants that unless C will be there. But some of the desires in McDaniel and Bradley's examples have conditions that are only tacit, and they posit a hidden 'condition' parameter underlying their ascriptions.

Now we may add that, as a matter of fact, there are few if any genuinely *un*conditional desires. How often does someone desire that P *no matter what*? 'Do you want P even if your child should be murdered? Even if it were to bring on nuclear holocaust?' I suppose someone might desire permanent world peace at any price, but that would be irrational; there are other great goods that would be worth a little border skirmish somewhere. Likewise for an end to all disease, and any other grand outcome I can think of. That my own children be good people (McDaniel and Bradley 2008, 267)?[14] Not at the cost of nuclear disaster.—Still, there might *be* such a monomaniacal desire, irrational or not.[15]

So far, Armstrong's analysis of desire is essentially unaffected, for he may allow that some desires have conditional content. A state of agent S may be apt for the production of information-sensitively ($A \supset C$)-seeking behaviour and with $A \supset C$ as a counterfactual cluster point. (Alternatively, we could try to parse conditional desire, *à la* conditional bets, as a desire that the agent would have assuming the condition holds, but that is not an option for us, since Armstrong and I are talking about actual mental states that the agent is in as we speak.) But it isn't quite that simple, because the conditionality in a conditional desire is not truth-functional, else my desire for lunch (psst: so long as there hasn't been a nuclear holocaust) would automatically be satisfied by a nuclear holocaust.

---

[13] See Lycan (2012); Fara (2013). Fara's term follows Kent Bach's 'Do Belief Reports Report Beliefs?' (Bach 1997), but the problem she and I have raised for desire is much worse than that suggested by Bach for belief.

[14] See also Persson (2005).

[15] McDaniel and Bradley allow for unconditional desires a bit artificially, by letting such desires be 'conditional on' the necessary truths. In that, they follow standard practice in dyadic deontic logic, of representing unconditional obligation as obligation conditional on tautology.

So, can one desire the truth of a subjunctive conditional? Is there such a thing as ($A \mathbin{\Box\!\!\rightarrow} C$)-seeking behaviour and with $A \mathbin{\Box\!\!\rightarrow} C$ as a counterfactual cluster point? That sounds odd, but Armstrong has a model for it, because as before, he maintains that counterfactual truths must have categorical bases as truthmakers. So perhaps S's conditional desire aims at that truthmaker. But the tacit conditions on a desire are going to be very complex (try some 'What if's on yourself); the conditional objective will look more like $[(A \,\&\, B \,\&\, {\sim}D \,\&\, {\sim}E \,\&\, F \,\&\ldots) > C]$, and who knows what the truthmaker for that would be?

7. Conditional Desires

It isn't even that simple either, for a conditional desire is actually not a desire for the truth of a conditional. For one thing, tacit conditions or not, my desire is intuitively for what a standard ascription would say it is for. I do desire a beer, and world peace (and I want them *now*!), not just the truth of the conditional 'If there is a pub within walking distance and if the beer isn't tainted with anthrax and if it won't cause a nuclear holocaust and....' (Of course, I tacitly believe that all the tacit conditions are satisfied.) Also, logically, conditional desire works like conditional obligation. The Modus Ponens inference is felt to be valid: 'B wants to go out for a beer, if C is not there; C will not be there; therefore B wants to go out for a beer'; but the corresponding formal inference from desire with conditional content is invalid:

$D({\sim}C \rightarrow B)$

${\sim}C$

$\therefore D(B)$

Cf. the deontic:

$O({\sim}C \rightarrow B)$

${\sim}C$

∴ O(B).

And, predictably, desire is subject to puzzles parallel to the 'paradoxes' of deontic logic, most notably Chisholm's Contrary-to-Duty paradox.[16]

I deal with those issues semantically by offering a semantics for conditional desire inspired by Fred Feldman's solution to the deontic paradoxes:[17] S desires that P given the condition C iff some C & P world is better by S's lights than any C & ~P world. (We may think of 'better by S's lights' as some version of preferring.) You can check for yourself that that analysis blocks both the foregoing formal inference and the Contrary-to-Duty paradox.[18]

What does all that mean for Armstrong, who has no truck with semantics or logic? The key points are that (1) with rare obsessive exceptions, every desire is a conditional desire, and (2) even though we speak of the 'desire that P', a conditional desire is not straightforwardly a propositional attitude. Yet Armstrong's causal formula does cast desire as propositional, since a desire's content is determined by its defining objective.

Nonetheless, a conditional desire is a relation between propositions. McDaniel and Bradley take the relation as primitive; I offer an analysis, but my analysis too makes conditional desire such a relation. Therefore, Armstrong has the option of bringing back his nested-conditional strategy: assuming all the tacit conditions hold, the causal-counterfactual constituents of the desire do also. It remains true to say that S desires that P, because in the context, S and we tacitly agree that the desire's tacit conditional antecedents are true. As regards S's mental state, this amounts to saying that S believes each of the antecedents. Which, by hypothesis, S does.

Of course, Armstrong's model is not friendly to tacit propositional attitudes, because for him a belief or desire or intention is a determinate mental cause, identical with a particular brain state. He distinguishes between attitudes that are causally active at a given time and those that are not. For beliefs, he offers the analogy of information stored in a computer. 'It may be causally inactive: simply stored in the memory-banks. Or it may be causally active: either playing a part in producing a certain output on the tape, or else acting upon and changing other internal states of the computer' (*MTM*, 344). But this does not help much, because truly tacit beliefs are not stored anywhere. Perhaps Armstrong can go nested-counterfactual again, and say that were the tacit (non-)issues to be called to S's attention, S would immediately acquire the relevant occurrent beliefs.[19]

---

[16] See Chisholm (1963). Example for the case of desires: (i-d) You want to go for a beer. (ii-d) You want to tell your partner you're going out, if you are going for a beer. (iii-d) You want not to tell your partner you're going out, if you are not going to go for a beer. (iv-d) You will not [in fact] go for a beer. (i-d)–(iv-d) are mutually consistent, but (i-d) and (ii-d) seem together to entail that you want to tell your partner you're going out, while (iii-d) and (iv-d) entail that you want not to. The four statements cannot be represented as consistent by any standard logic of conditionals and sentence operators. (Qualification: It's assumed, not that there can't be contradictory desires, but that no contradictory desire figures in the present setup.)

[17] See Feldman (1986), following Lewis (1973) and van Fraassen (1972). The present analysis is revised from my 2012 article and defended in 'In What Sense Is Desire a Propositional Attitude?', rejected by *Mind*.

[18] It does not sustain the MP inference. But I do not believe that the MP inference is formally valid, though in a context it is a perfectly proper inference. The inference is usually made in the first person, and its conclusion is solidly assertible, but only because it is tacitly assumed that the tacit conditions hold. When the inference is made in the third person, it is likewise felt to be correct just when the tacit assumptions are firmly in place.

8. Conclusion

Armstrong's account of conation has held up astonishingly well under fifty years of ever more sophisticated research. That is quite a tribute to a philosopher who has already deserved and received many more general tributes. I only wish he were here to add new observations and insights on the topic.

9. Afterword on Intentional Content

In the 1960s, for reasons I do not know, intentionality was not much talked about. Concern for intentionality was being kept alive primarily by Roderick Chisholm, and even Chisholm was focusing on grammatical and logical features rather than metaphysics, as then did Elizabeth Anscombe.[20] Armstrong mentions intentionality several times but never dwells on it as such (save in Section V of Chapter 11, regarding perception, on which more shortly). It would be another ten years before issues of mental referring, and the psychosemantic question in particular, would become current.

Armstrong went on to address a couple of them in his 1984 *Profiles* volume reply to Steve Stich.[21] First, his reluctant thoughts on content externalism; then, a remarkable psychosemantic idea to which attention should be paid.

Stich's main concern was to confront Armstrong with, or make Armstrong confront, the 'narrow'/'wide' distinction and the problem of relating wide content to psychological explanation. Of course, Armstrong knew of Putnam's Twin

---

[19] On such issues, see my 'Tacit Belief', reprinted as Chapter 3 of Lycan (1988). I am reminded that in conversation *c.*1983, Armstrong expressed scepticism about the notion; maybe there are no truly tacit beliefs that are states or properties of people. But I think he needs it here.

[20] See Chisholm (1955–56); Anscombe (1965).

[21] 'To Stich', in Armstrong (1984, 237–43), in response to Stich's 'Armstrong on Belief', Armstrong (1984, 121–38). Stich had already published a splendid paper defending content externalism, 'Autonomous Psychology and the Belief-Desire Thesis' (Stich 1978).

[22] In my 1983 seminar in T&M, Kim Sterelny had verbally hammered him with Twin Earth (Putnam 1975); Michael Devitt, then a recent Putnam student, was also present and (shocker coming now!) vocal. But Armstrong then seemed to think the point was unimportant.

Earth,[22] but he had never come clean on the matter. Stich rightly complained that Armstrong's purely causal analysis of the propositional attitudes seemed to leave him unable to account for wide intentional content (i.e., for intentional content). In response to Stich's typical examples in which two people just alike in their head contents have parallel beliefs about different individuals or kinds, Armstrong 'grant[s] Stich that there is *a* sense in which' the two people's beliefs differ (Armstrong 1984, 238, italics original), but he maintains that 'there is another sense, a sense already present in ordinary thought', in which the people '*have the very same belief*' (Armstrong 1984, 238–9, italics original). That is, he simply insists that there is narrow content. In that, of course he is not alone, but he surely underestimated the difficulty of specifying a narrow attitude content while preserving its status as (propositional) *content*, especially since he thinks that a narrow sort of content is an 'ordinary' sort. (He further suggests, not implausibly, that the wide intentional differences may be explained as causal differences.)

In *Belief, Truth and Knowledge*, Armstrong had tried to explain the intentionality of concepts in terms of discriminative capacities. Stich challenges him on that by suggesting that the concept 'cabbage' refers to the sadly disjunctive class of cabbages, cabbage imitators and hallucinations a person's imperfect cabbage detector would in fact pick out, rather than to all and only cabbages. Not then wishing to make the now common appeal to teleology,[23] Armstrong offers a truly original psychosemantic suggestion:

> We have, or we can have, purposes which will be satisfied if and only if the object is a cabbage... [He means purposes, not as teleofunctions, but as mental states as analyzed above.]
>
> What I am trying to do here is to fix the intentionality of the selective capacity by reference to the intentionality of the purposes of which it is the servant: in this case, purposes involving cabbages...
>
> I do think that a good deal of progress is made by linking the problem of intentionality of our simple, or simpler, concepts with the problem of the intentionality of purpose. The linkage respects the package-deal nature of the fundamental mental concepts. (Armstrong 1984, 241)

Wow. Contrast the more usual way of understanding intentionality derivatively,[24] namely, to give a psychosemantics for perception (the easy case for a causal-cum-teleological psychosemanticist), understand the intentionality of beliefs as deriving from that of perception, and only then derive the intentionality of desire and other attitudes from that of beliefs (or much more likely, just ignore all other kinds of intentional states). But it is a great virtue of Armstrong's scheme—to my knowledge, never mentioned in the literature[25]—that he gets the intentionality of desire pretty easily in terms of the end or objective.

---

[23] Later, in the Preface to the long overdue paperback edition, penned in 1992, he acknowledges that it 'would strengthen' the Causal theory to add a teleological element; in particular, it helps to ask what a sensation 'is *for*' (Armstrong 1993, xviii, italics in original).

[24] For example, as in Brian Loar's *Mind and Meaning* (Loar 1981).

Moreover, the idea was not new in response to Stich, but had been quietly defended in 1968, in regard to perception in infants:

> We can now see that the problem of what makes a certain mental event the perception that a certain situation obtains—the problem of the intentionality of perceptions—is bound up with the intentionality of the will. The notion of the intentional object of a perception is bound up with the notion of objective. For suppose that I perceive that there is a blue block in front of me. I might manifest my perception by putting the block into a box marked 'blue' [because that is my purpose in regard to what I see] . . .
>
> [Once we] describe objectives of discriminatory behaviour in the ordinary way, . . . then we can say that perceptions have the intentional objects that we ordinarily say they have. (*MTM*, 262–3)

But why should we acquiesce in what we ordinarily say? Armstrong goes on to explain that in terms of 'the *causal circuit* involved in purposive behaviour that has as its objective some change in non-relational or relational properties of the object perceived' (*MTM*, 263, italics in original); 'the thing perceived acts on the perceiver, who in turn brings about some change in the non-relational or relational properties of the situation, and this new situation in turn acts on the perceiver to inhibit further action' (*MTM*, 263).

And moreover: 'The [*full* causal] circuit goes further out into the world: it goes as far as the blue block, although it goes no further. The blue block plays an essential role in the complex back-and-forth causal transaction, or, at any rate, the blue surface of the block does' (*MTM*, 264). The basis of content externalism, despite Armstrong's distaste for same.

References

---

[25] At the 2018 conference, Dave Chalmers reminded me that a psychosemantics based on conative ends or objectives might be extracted from some 'enactive' theory of intentionality. James Franklin contended that to allow a conative or at least pragmatic element even into psychosemantics was a self-departure from Armstrong's naturalistic realism—a point as much worth investigating as is the idea itself. Thanks to Chalmers, Franklin, Frank Jackson, and many other conference participants for very useful discussion.

Anscombe, G. E. M. (1965) 'The Intentionality of Sensation: A Grammatical feature', in *Analytical Philosophy: First Series*, edited by R. J. Butler, Oxford: Basil Blackwell, 158–80.

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1980) 'The Causal Theory of the Mind', in *The Nature of Mind and Other Essays*, St Lucia: University of Queensland Press, 16–31.

Armstrong, D. M. (1984) *D. M. Armstrong: Profiles*, vol. 4, edited by Radu J. Bogdan, Dordrecht: D. Reidel.

Armstrong, D. M. (1993) 'Preface', in *A Materialist Theory of the Mind*, 2nd edition, London: Routledge, xiii–xxiii.

Bach, Kent (1997) 'Do Belief Reports Report Beliefs?', *Pacific Philosophical Quarterly*, 78: 215–41.

Braithwaite, R. B. (1953) *Scientific Explanation: A Study of the Function of Theory, Probability and Law in Science*, Cambridge: Cambridge University Press.

Bratman, Michael (1987) *Intention, Plans, and Practical Reason*, Cambridge, MA: Harvard University Press.

Chisholm, R. M. (1955–56) 'Sentences about Believing', *Proceedings of the Aristotelian Society*, 56: 125–48.

Chisholm, R. M. (1963) 'Contrary-to-Duty Imperatives and Deontic Logic', *Analysis*, 24: 33–6.

Davidson, Donald (1963) 'Actions, Reasons and Causes', *Journal of Philosophy*, 60: 685–700.

Dennett, Daniel C. (1971) 'Intentional Systems', *Journal of Philosophy*, 68: 87–106.

Fara, Delia Graff (2013) 'Specifying Desires', *Noûs*, 47: 250–72.

Feldman, Fred (1986) *Doing the Best We Can*, Dordrecht: D. Reidel.

Lewis, David (1973) *Counterfactuals*, Cambridge, MA: Harvard University Press.

Loar, Brian (1981) *Mind and Meaning*, Cambridge: Cambridge University Press.

Lycan, William G. (1988) *Judgement and Justification*, Cambridge: Cambridge University Press.

Lycan, William G. (2012) 'Desire Considered as a Propositional Attitude', *Philosophical Perspectives*, 26: 201–15.

Martin, Adrienne (2014) *How We Hope: A Moral Psychology*, Princeton, NJ: Princeton University Press.

McDaniel, Kris and Ben Bradley (2008) 'Desires', *Mind*, 117(466): 267–302.

Nagel, Thomas (1970) 'Armstrong on the Mind', *The Philosophical Review*, 79(3): 394–403.

Persson, Ingmar (2005) *The Retreat of Reason: A Dilemma in the Philosophy of Life*, Oxford: Oxford University Press.

Putnam, Hilary (1975) 'The Meaning of "Meaning,"' in *Language, Mind, and Knowledge,* vol. VII, *Minnesota Studies in the Philosophy of Science*, edited by K. Gunderson, Minneapolis, MN: University of Minnesota Press, 131–93.

Schueler, G. F. (1995) *Desire: Its Role in Practical Reason and the Explanation of Action*, Cambridge, MA: MIT Press.

Stich, Stephen (1978) 'Autonomous Psychology and the Belief-Desire Thesis', *Monist*, 61(4): 573–91.

Stich, Stephen (1984) 'Armstrong on Belief', in Armstrong (1984), 121–38.

Van Fraassen, Bas (1972) 'The Logic of Conditional Obligation', *Journal of Philosophical Logic*, 1: 417–38.

12

Causal Conditionalism

David Braddon-Mitchell

1. Introduction

Analytic functionalism of some kind has seemed to me the least unacceptable of the various positions in the metaphysics of mind, and I've spent much of my career defending and ultimately modifying it in the light of objections. In this, I owe a great debt to Armstrong, who took the insights of Lewis and Smart and others and produced what remains perhaps the most elaborate and detailed defence of what he called the 'causal theory of the mind' in a range of domains. One of the two most persuasive objections to the view that has emerged is what has become known as the 'conceivability argument' in various forms: from Campbell's 'imitation man' to Chalmers' zombies (Campbell 1970; Chalmers 1996). This chapter is yet another response to that argument. In one way, it is a modification of an earlier reply of mine to that argument (but not, I think, a replacement—which one works better dialectically will depend upon the opponent, as we'll see). In another way, it's dipping back to Armstrong's thought, and using another resource—some remarks of his about causation—to forge a better defence of his view without the modifications involved in my own treatment.

2. The Conceivability Argument

The conceivability argument against materialism is now very familiar to most with an interest in the philosophy of mind. It starts by asking you to imagine a world which contains exact physical duplicates of the conscious beings in our world. For simplicity, let's suppose the entire world is a complete minimal physical duplicate of our world.[1] Then you are asked to imagine that the physical duplicates therein have no conscious states—they are, in Chalmers' now famous phrase, 'philosophical zombies'. There's nothing going on inside them. The next step is to argue that this kind of imagination amounts to conceivability. We can conceive of philosophical zombies. Then the thought is that conceivability, at least of a certain

---

[1] A physical duplicate is minimal just in case it duplicates all of the physical properties of our world, and does not include any additional properties not found in our world.

David Braddon-Mitchell, *Causal Conditionalism* In: *Armstrong's Materialist Theory of Mind.* Edited by: Peter R. Anstey and David Braddon-Mitchell, Oxford University Press. © David Braddon-Mitchell 2022.
DOI: 10.1093/oso/9780192843722.003.0013

idealized kind, is a good (and perhaps ultimately the best and only) guide to possibility, and thus that these zombies are possible. So, there is a possible world physically exactly like ours and yet there is no consciousness in that world. What, then, accounts for consciousness in our world? It cannot be anything physical, because this duplicate world is exactly like ours physically and yet contains no consciousness. So, consciousness in our world must have a non-physical basis, and thus materialism is false.

There is, of course, a huge number of attempted replies. Many concentrate on denying that even ideal conceivability is a good guide to possibility, especially metaphysical possibility. For what it's worth, I don't think this line of response works, but this is not the place to pursue that. Instead, we'll see what other lines of response might be effective—and dialectically at least this is critical, since I am far from alone in being unpersuaded by this move.

Analytic functionalists have a straightforward response. If you are Armstrong (or any other analytic functionalist), you think that analysis is involved in the first stage of a theory of the mind. What mental states are, is states apt for causing certain kinds of behaviour. This is not an a posteriori discovery. Materialism about the actual world, however, is an a posteriori matter. It's the empirical discovery that the states apt to cause certain behaviours are material or physical states.

So, what of the alleged zombie world? Can we conceive of it? Certainly not. For it is a priori that whatever plays certain causal roles is a mental state. Our world contains states that play certain causal roles. So, it follows from what we know a priori that it contains mental states. The zombie world duplicates those causal roles, so the idea that it's a zombie world is inconceivable: it is ruled out a priori and is thus impossible. Whatever this quasi-visual imagining people say they can do must not be conceiving in the strict sense at all, far less ideal conceiving. Problem solved; materialism vindicated.

3. Some Caveats and an Alternative

If you think this sounds too easy, then I agree. There is nothing wrong with the logic of the reply, of course. If it really is true that it's a priori that mental states are whatever plays certain casual roles, the response is sound. But the remarkable facility of the reply instead draws attention to what was perhaps always a bit facile about some versions of analytic functionalism (versions I've called procrustean functionalism elsewhere). If it's an a priori conceptual truth, why are so many unaware of it, and indeed why do so many philosophers deny it? The claim that it is a conceptual truth that whatever plays certain functional roles is a mental state does not seem to be the kind of difficult, hidden, a priori truth that might, perhaps, take a mathematician her entire career to establish, and still result in her not being believed. It's supposed to be fairly straightforwardly what we *mean* by mental state language. There are persuasive arguments that this kind of a priori truth is embedded in our practice tacitly, as part of a very complex theory of the mental, and perhaps plays less of a role in conscious deliberation, and thus it's easy for people to deny it without seeming contradiction. But this requires you to hold, nonetheless, that anyone who thinks zombies are possible is simply conceptually incompetent, which is a strong claim indeed. So, the worry is that it is all too simple, and the idea that it's a relatively straightforward fact of meaning that functionalism is true, too implausible. So, if there could be a defence of functionalism or the causal theory against the conceivability argument that did not rely on it, we would be better off.

That's what I offered in 2003. In 'Qualia and Analytical Conditionals' (Braddon-Mitchell 2003), I argued that it's not straightforwardly analytic that whatever plays certain causal roles is a mental state. Rather, something more complicated is analytic. This more complicated thing is a conditional: roughly 'If the actual world contains "spooky"[2] states, then mental states are those spooky states and necessarily so, or else mental states are whatever states play certain causal roles.'

This kind of conditional better explains people's behaviour and the patterns of assertion. If you are a dualist, then you think that the actual world contains spooky states. It follows from the conditional that they are necessary for mental states, and thus that a minimal physical duplicate of the actual world fails to contain them, and so has no mentality and is a zombie world. On the other hand, if you are a *completely* convinced physicalist, you should think that the actual world contains no spooky states, and then the conditional tells you that states playing certain functional roles are sufficient for mentality. Thus, you will judge that a minimal physical duplicate world must contain mental states and that zombies are impossible. Why then does even the physicalist take it to be in some sense conceivable that zombies are possible? Because it's not a priori that zombies are impossible. What's a priori is that if the actual world contains no spooky states, then zombies are impossible. And a physicalist who is minimally rational by Bayesian standards gives non-zero credence to the actual world containing spooky states. So, the physicalist confuses, perhaps, a non-zero credence in it being right to judge that zombies are possible, with an ideal positive conceiving of their possibility.

I remain largely persuaded by this position; but there are two influential objections and in this chapter I offer an alternative view which has something of the spirit of the original view, but which avoids these two objections, and is a cleaner reply for someone who would like a reply which nevertheless leaves analytic functionalism intact. The new view, then, will be preferred by those who are moved by the two objections I'll describe, and who are looking for a reply that preserves analytic functionalism. It's thus a view even more directly in the spirit of Armstrong and Australian functionalism, as it preserves unaltered the causal theory, and—as I shall spell out—takes inspiration from remarks of Armstrong on causation.

---

[2] This was not intended as a term of derogation, rather as a shorthand for something like 'states of a non-physical nature of which the doctrine of relegation is true, i.e., that their essential nature is revealed, non causally, in experience'.

4. Two Objections to the Conditional View

The first objection to the conditional view is that the concept of any mental state, or any type of mental state is unlikely to be so complicated as in my original 2003 paper, or in any account which defends a similar view. No-one introspects on their concept of <qualia>, the objection runs, and sees a conditional whose antecedent mentions dualistic states, whose consequent mentions dualistic theories of the mind, and which contains an 'else' clause that asserts functionalism.

Of course, the background to the view is an account of concepts, according to which conceptual competence is revealed in patterns of judgement and behaviour. Few concepts are introspectible, and to the extent that they are, our introspected theories about our own concepts are fallible (see, e.g., Ravenscroft and Stich 1994). So perhaps we should not be overly concerned about such an objection. But, of course, on some accounts of concepts, the concern remains, which means that there are dialectical reasons for taking it seriously and formulating a reply to the conceivability argument which does not fall foul of the objection.

A second objection to the view is that the conditional cannot be a priori, not because it's too complicated, but because it depends on something a posteriori. This is the worry that Torin Alter (Alter 2007) has that the conditional is clearly motivated by something a posteriori—the knowledge that we are in fact conscious. Absent that, he argues, the idea that *come what may* we are conscious, regardless of how the actual world is, would have no appeal. He thinks that because we know a posteriori that we are conscious, it might be true that we would adjust our beliefs to accommodate physicalism if our world turned out to be physical. But this is no conceptual truth: it's just adjusting our views in the face of a posteriori knowledge.

Again, in accommodating this objection, I don't endorse it. In my view, it follows from possessing the full suite of concepts that we do know that we are conscious—thus, it follows from our concepts and reason alone that we are conscious, enough to make it a priori that we are conscious on some, though not all, conceptions of what it takes to be a priori.[3] This chapter is not the place to rehearse this reply, though, since the objection itself is a motivating factor in seeking an additional, more Armstrong-friendly, response.

---

[3] For a survey of what many of these are, see Boghossian and Peacocke (2000).

5. The New Theory

It's time to reveal the alternative. The key idea here is to retain analytic functionalism (or the casual theory, as Armstrong preferred to call it) and look for the source of the conditional behaviour elsewhere.

According to analytic functionalism, a (minimal) functional duplicate of a world is a mental duplicate of a world. Duplicate the causal functional roles and you have duplicated the mental. What is the role of the 'minimal' in brackets above? Well, the thought is that if you, having duplicated the causal functional roles of the actual world, proceed to add other things, your causal functional duplicate may no longer be a mental duplicate. These other things might include further functional roles, non-physical raw consciousness (for versions of functionalism which make it a view about what is *sufficient* for consciousness, but allows dualistic consciousness to be logically possible albeit not actual), and so on.

So, functionalism does not admit of the possibility that a minimal causal functional duplicate could fail to be a mental duplicate, and thus it cannot allow for the possibilities of philosophical zombies, where zombies are thought to be causal functional duplicates which lack the mentality of their originals.

But, of course, the conceivability argument is not expressed in terms of causal-functional duplicates. Instead, it is expressed in terms of *physical* duplicates. This has never seemed to concern us because of the close connections between functionalism and physicalism, but perhaps it should have.

It's fairly well known that functionalism does not make materialism a *necessary* truth. One of its benefits is that it allows for the logical possibility of worlds where the causal functional roles are played by non-material things. The benefit is supposed to lie in the fact that we can conceive coherently of dualistic worlds consistent with our concepts of the mental, but we find that we are in a physical world when we discover a posteriori that the things playing those roles are physical. This is how Armstrong conceived matters. The causal theory is a piece of analysis, the discovery that what plays the roles is physical is delivered to us by science, and thus is contingent. If there are worlds where immaterial things cause behaviour, these are not materialist worlds.

This understanding, though, has never (to my knowledge) made anyone doubt that functionalism makes zombies impossible, even when we conceive of a zombie world not directly as a functional duplicate, but rather as physical or material duplicate. Why? Because it's thought that being a material duplicate *entails* being a causal functional duplicate. So, the view works this way: imagine a minimal material duplicate of the actual world. Such a duplicate is also a minimal functional duplicate—all the material machinery of causality is duplicated. Since it's a functional duplicate, it's a mental duplicate, so it cannot be a zombie world if the actual world is a mental world.

Enough scene-setting: what has been neglected here is the claim that a material duplicate is a causal functional duplicate. What if it is an epistemically open question whether causality itself is a material phenomenon? I hasten to add I think it *is* a material phenomenon, but not everyone seems to agree, as I'll explain later. If it's a conceptual possibility that causality is not a material phenomenon (even though associated with materiality in the actual world), then the conceivability of zombies is explained by the conceivability of a material copy of the actual world which has no causing: it's just one thing after another without true causal connections. Of course, on some Humean accounts of causality, that is enough if a pattern is thereby formed: but these are by no means the only views. And all it takes for conceptual possibility is for there to be coherent views on which such a mere material duplicate lacks causation.

If there can be minimal material duplicates of the actual world which lack causation, then these duplicates are *not* causal functional duplicates. And if they are not causal functional duplicates, indeed, worse, if they lack causal functional roles, they lack mentality. So they are zombie worlds. So not only then does analytic functionalism admit the conceivability of material duplicates that are zombies, it fairly directly admits the possibility of them, if those material duplicates lack causation. And it recovers the conditionality of my original 2003 proposal (Braddon-Mitchell 2003) (and that of Hawthorne (2002)) without it being built into the concept of the mental. For if causation is in fact immaterial (as I doubt), then we should judge that zombies are possible. But if we were to come to believe that causality is a material thing, and material duplicates do in fact entail causal duplicates, we should judge that zombies are impossible. As before, if you find yourself located in the 'else' clause of this conditional, thinking that causation is material, the conceivability of zombies is explained as a shadow of your acceptance that you might be mistaken about the material nature of causality. That ends a rough explication of the view: now to unpack some of its components, and consider some replies. I'll start with this idea that causation might or might not be a material matter.

6. Causation Has a Hidden Nature

For this idea to work, we need to think that for all we know about the concept of causation, it might (or might not) be a material thing. So, what kind of concept of causation might have that feature? It's in Armstrong that I found the thought that sparked this. Here is what he says in *Bodily Sensations*:

> The concept of pressure involves the concept of causation, of one thing making another happen. But can we have immediate perception of causal relationship? If we cannot, pressure cannot be immediately perceived. The question is a difficult one, and deserves a more extended discussion than I am able to give it. But I am inclined to say that there is no objection to the immediate perception of causal connection. (And, at the same time, I do not think that this necessarily contradicts a Humean or semi-Humean analysis of the nature of causation.) It is simply one of the peculiar features of tactual perception that it gives us an immediate awareness of objects making things happen to our body, and our body making things happen to objects. (Armstrong 1962, 23)

The thought here is that the concept of causation is something that we get from our experience of the world impacting on us. Whatever that thing is which underlies the impact (and perhaps the possibility of our impacting the world) is what causation is. The idea that Armstrong thought that this understanding was 'topic neutral' (the term J. J. C. Smart used to indicate an analysis that is silent about the underlying nature of the thing analysed) is indicated by his parenthetical comment that 'I do not think that this necessarily contradicts a Humean or semi-Humean analysis of the nature of causation.' So what we have from our concept of causality is something like an arrow asking us to look elsewhere; it's silent about a more detailed account (or analysis—I'm not sure that Armstrong intended 'analysis' here to be conceptual) of the nature of causation. If it doesn't 'necessarily' rule out a Humean analysis, then presumably it must not rule out most of the accounts of causation that we find in the philosophical literature.

A striking feature of accounts of causation in the literature, however, is that they do not all assume that causation is a physical phenomenon. Sometimes this is because they take it to be a logical relation; but I'm interested here in the thought that it might be a metaphysically substantive relation that is over and above how things are physically, at least in as much as it fails to supervene on intrinsic physical nature. Let's call such a view—which I do not hold, but which I take to be conceivable—the view that causation is not physical. The conceivability is what we need to undergird the proposal preliminarily sketched in the previous paragraph.

So, Armstrong's suggestion—that causation may be whatever it is that is going on in paradigm cases of causation—is consistent with all sorts of proposals about what it is that's going on. What I need is that one such option is that it is not physical, thus allowing physical duplicates that are not causal ones.

Fortunately, they are not far to seek. Indeed, something like this idea is central to a range of responses to a problem in the philosophy of time, about which it will be helpful to provide a brief excursus.

The 'growing block' view of time is a view which includes a four-dimensional past—a 'block' of the kind familiar to the B theorists—with an A theoretic component. The block grows over time. The past exists in concrete space-time, but the future does not exist. As time passes, new slices are added to the 'salami' of time. It's a view designed to combine solutions to the problem of what makes true claims about the past (the past does, easy!) with openness about the future (it's empty, nothing there yet!).

A fairly influential objection to this emerged in the mid-2000s (Bourne 2002; Braddon-Mitchell 2004), which has become known as the epistemic objection. The idea, roughly, is that if this is true, the past is all a physical reality, but—according to the growing block—there is only one objective present, the last slice of this objective physical reality. Here's one way to characterize the objection (there are in fact a few, but this doesn't matter for our purposes). In 2004, say, all of history is physically real, but only one moment in 2004 is the present. In all of that past reality there are countless agents with mental states having all the same thoughts they had back when they were in the present. One of the thoughts many are having is that they inhabit the present. Only the agents in the present are right about this. I think I'm in the present, but I see (on accepting this view) that most agents who think this are in the past, so I'm probably mistaken. So, I'm probably not in the present, or indeed anywhere near it. So, to simplify a lot, by *reductio*, the growing block view is mistaken.

Advocates of this view have many responses, but one response which many have adopted (Forbes 2016; Forrest 2004) is that this doesn't follow because agents in the past are all philosophical zombies. Why? Because while they are physical duplicates of conscious agents, *nothing is going on in the past.* There is no causation in the past, because causation is a special fizz which happens only in the border between being and non-being: at the last slice of the salami, the objective present. Thus, on functionalist views about mentality, while the past may have brains with physical states that seem to be apt for consciousness, those physical states are not playing any causal roles, and thus there's no consciousness. So, there is one set of theorists happy with the idea that consistent with our shared concept of causation, it might be that causation outstrips the physical.

Here is another. There is a model of a possible universe due to the cosmologist Thomas Gold, according to which the universe is symmetrical: it starts with a big bang, expands and entropy increases, and it then contracts with decreasing entropy until there is a singularity at the other end (Gold 1962). This raises a puzzle: what is the arrow of time? If the arrow of time follows increasing entropy, as many think (see Price 1997, for a defence of that view in this context), there is not an end and a beginning, there are two beginnings each pointing to the 'end' state in the middle of this universe. Many (Maudlin 2007, for example) take the view that this would just be a weird world in which there is a single objective direction of time and causality, and just a whole bunch of physical stuff in one half of the universe arranged in a way contrary to the true direction of time and causation: it's a world in which, while half of the world contains the same physical properties and relations, there's something else that we cannot learn from physics alone missing in the other half.[4] Thus, for Maudlin (2007), while there may be beings that look from the outside like agents with thoughts and experiences in the half of the world going the 'wrong' way, they are philosophical zombies.

There are, of course, other views about causality which take it to be something over and above the (straightforwardly) physical. One pertinent possibility is that the Armstrong-Tooley view is one of them (Armstrong 1983; Tooley 1987); instances of contingently holding relations of necessitation sound like something neither physical nor logical, and as such might be included in my catalogue of existence proofs of non-physical accounts of causation. But I don't plan to add more: enough that it seems open enough.

7. From Hidden Causation to the Conceivability of Zombies

Your present author disagrees with both these claims. But the point is that many intelligent, smart philosophers do not. They think two crucial things:

1. That causation is something over and above the properties mentioned in physics and their arrangement, and
2. That causation in this sense is necessary for consciousness. Consciousness is created by the real causal behaviours, not by patterns.

The extent to which I agree with them is that I don't think the concept of causation settles by itself that they are wrong and that folk like me are right. Furthermore, if it turned out that, say, Forrest was right about how things work in the actual world, I would be sufficiently impressed that I would take this not to be a mere peculiarity about the actual world, but to tell us something about the essential nature of causality. As Armstrong might have thought, causation is whatever it is we are in contact with in certain paradigm situations.

So, once again we have a conditional: the idea is that Humean or (for example) energy-transfer conceptions of causation are deflationary. They'll do for causality if the world offers up nothing else (as I, for one, expect). So, we get:

Conditional Causation

IF the world contains non-physical generative causation of some kind

THEN that's what causation is and necessarily so

ELSE it is sufficient for causation that we have a Humean or other deflationary substitute.

---

[4] Though Maudlin sometimes denies this, it's unclear whether it's physics alone or physics plus philosophy at work.

How does this finally solve the problem?

Well, the thought is that for the analytic functionalist, it's not a priori that physical duplicates of the actual world are mental duplicates, even though analytic functionalism is a priori. For we cannot rule out a priori that causality is something over and above the physical properties in the world and their distribution. So maybe there is a world which is a physical duplicate of ours, and yet is not a causal duplicate. Analytic functionalism requires that a world be a causal duplicate for it to be a mental duplicate: so, it's open that some physical duplicate world is a zombie world.

That this should be conceivable is consistent with holding that physicalism about the mind is true and that zombies are impossible. For if you hold that as a matter of (empirical or metaphysical but not a priori) fact that causation in the actual world is a purely physical matter, according to the conditional concept of causation, then being a physical duplicate is sufficient for being a causal duplicate. Since it is a priori that causal duplicates are mental duplicates according to analytic functionalism, then no possible minimal physical duplicate can fail to be a mental duplicate, and thus zombies are impossible.

The argument goes thus:

1. a priori premise: Causation is whatever we are in touch with via paradigm instances.
2. a priori conditional as above: If we are in touch with something non-physical underlying paradigm instances, this non-physical nature is necessary for causation. Otherwise, the physical things we are in touch with are sufficient.
3. a posteriori premise of physical causation: we are in touch with purely physical phenomena in causation.
4. Lemma*:* from 1, 2, and 3: physical duplication is sufficient for causal duplication.
5. a priori from definition of analytic functionalism: causal duplication is sufficient for functional duplication.
6. a priori from definition of analytic functionalism: functional duplication is sufficient for mental duplication.
7. Lemma from 4, 5, and 6: physical duplication is sufficient for mental duplication.
8. a priori from definition of zombies: Zombies are physical duplicates that are not mental duplicates.
9. Conclusion from 7, 8: Zombies are impossible.

Thus, the analytic functionalist can hold that zombies are impossible, via this argument. But the argument depends on an a posteriori premise: 3. So while we might judge that zombies are impossible, our conviction that this is so is not wholly a priori. While we might take ourselves to have good reason to endorse the impossibility of zombies, as rational beings, we give some credence above zero to a posteriori claims which are inconsistent with what we take to be most likely. And we see that, given our conditional concept of causation, if we thought that as a matter of fact the phenomena we are in touch with via paradigm instances of causation were not physical, as many philosophers believe, we would be in a position to make the following argument:

1.* a priori premise: Causation is whatever we are in touch with via paradigm instances.
2.* a priori conditional as above: If we are in touch with something non-physical underlying paradigm instances, this non-physical nature is necessary for causation. Otherwise, the physical things we are in touch with are sufficient.
3.* a posteriori premise of non-physical causation: We are in touch with non-physical phenomena in causation.
4.* Lemma*:* from 1, 2, and 3: Some non-physical phenomena are necessary for genuine causation.
5.* Minimal physical duplicates are not causal duplicates.
6.* a priori from definition of analytic functionalism: Causal duplication is *necessary* for functional duplication.
7.* a priori from definition of analytic functionalism: Functional duplication is necessary for mental duplication.
8.* Lemma from 4, 5, and 6: Some minimal physical duplicates are not mental duplicates.
9.* a priori from definition of zombies: Zombies are physical duplicates that are not mental duplicates.
10.* Conclusion from 9: Zombies are possible.

Even given the a priori status of analytic functionalism, we cannot know a priori whether 3 or 3* is the case, and thus we do not know which of these arguments is sound. So, the thought is that the physicalist who thinks zombies are impossible in some sense, in conceiving of zombies is in fact doing something else: they are conceiving of the actual world being such that, given our concepts, it would be right to say that zombies were possible. Exactly how this view vindicates or fails to vindicate the idea that conceivability is a good guide to possibility is something that, on this view, is no different than in my 'Qualia and Analytical Conditionals' (Braddon-Mitchell 2003), so I will not say much here, except to note that the 'zombies' we are conceiving of in the two cases are not exactly the same thing: it's more that the physicalist sees that while they judge zombies to be impossible, they see that if the world were different, we should call things zombies that we do not now call zombies. So, it is not a straightforward matter of conceiving of something being possible that is not. But this subtlety cannot be expected to be reflected in our capacities to conceive of states of affairs.

If I am right about all this, then we have a reply to the conceivability argument for someone who thinks that analytic functionalism about the mind is a better candidate for an a priori theory of the mind than the conditional story, but who is prepared to accept that conditionalism about causation is true. So, it's time to turn to an alternative, and a possible reply.

8. An Alternative: Rigidified Analytic Functionalism

There is an alternative that will give us the same results without appealing to a conditional concept of causation. Instead of endorsing straightforward analytic functionalism, we adopt rigidifying analytic functionalism: a view I've elsewhere called empirical functionalism. According to this view, analytic functionalism is not a view about what is necessary and sufficient for mental states. Rather, it's a view about how to select samples whose internal, hidden natures are what is necessary for mentality. So, the idea is that analytic functionalism tells you that the actual, local things of which analytic functionalism is true are the paradigms (and indeed it's always true for anyone who adopts this view that the actual and local things of which analytic functionalism is true have mental states). We then analyse these actual local states, and find what their hidden nature is, and that is what it takes to have mental states. There might, for all we know, be counterfactual or merely distant things of which analytic functionalism is true, which do not have mental states because they lack the hidden natures of the actual and local things of which analytic functionalism is true.

This might be thought to neatly avoid the conceivability argument. We do not know what the hidden nature is of the actual local things of which analytic functionalism is true. For all we know, this hidden nature is dualistic. If it is dualistic, then, of course, there are physical duplicates which are not mental duplicates. If it is physical, then, of course, there aren't. We cannot tell which of these is the case a priori, so our capacities to conceive leave it open.

But it's not that straightforward. First, the view is unattractive in various respects, second, it does not do the dialectical work of solving a problem for analytic functionalists, and, third, to the extent that we can clean it up, it turns into the conditional qualia view and thus is not a distinct view.

The unattractiveness is spelt out in some detail in Braddon-Mitchell and Jackson (1999) but there are at least two concerns. One is that it is silent on what 'hidden nature' amounts to. There are countless things true of individual objects that behave the way analytic functionalism says they have to behave to count as mental. There is cognitive architecture, neural architecture, physical nature, evolutionary history, and (perhaps) underlying metaphysical nature.

Which of these is the thing we need to rigidify on? It seems some further constraint is called for. The second is what 'local' amounts to. If it's just humans, and you need to be neurally like us, then great apes and octopuses are ruled out. If all the intelligent seeming life on Earth is what counts as local, and we rigidify on what is in common among that, then we rule our nearest intelligently behaving neighbours in the galaxy, if any. If we include them, then we have no principled reason to restrict 'local' to anywhere in the universe: so, we are looking for what is in common between all actual things of which analytic functionalism is true. But then someone will complain that surely the things, which, by mere accident, failed to evolve intelligence cannot in principle be ruled out. So, soon we include all possible things of which functionalism is true. And there is no reason to suppose there is anything true of all of these other than that analytic functionalism is true. So, the view then would collapse to analytic functionalism.

The second worry is the dialectical one that even if the view were more attractive, it might solve the problem, but not for analytic functionalists. It is no small emendation to analytic functionalism, which is supposed to be the view that in causal roles we find the essence of the mental. Rigidifying functionalism says only that we use functionalism to locate the paradigms around us—functionalism plays no part in specifying their essence. In particular, it falls foul of the functionalist objection to what Galen Strawson (1994) thinks is a real possibility: something that in virtue of being intrinsically like a mind, while having no causal powers whatsoever, gets to be conscious.

The final concern is that if we were able to constrain the worries of the first objection in a way that would make the view do the work of avoiding the conceivability argument, these constraints would make it collapse into some version of the conditional qualia view. Here's why: suppose we solve the problem of what we should rigidify on by simplifying it into two basic choices. Either we find a non-physical nature which perhaps obeys revelation, or we do not. If we do find such a non-physical nature, we rigidify on it, and take it to be necessary. If we do not find it, we do not concern ourselves with all the different sorts of natures that might subserve the purpose of playing the causal roles, and we simply rigidify on the fact that the causal roles are played. But now we have the conditional qualia view, not a distinctive one. If there are actual spooky states, they are the qualia and necessarily so. If not, then functionalism is true.

9. An Objection: The Conceivability Argument Reconfigured

The current proposal allows the analytic functionalist to make sense of the idea that, consistent with their mental concepts, they can imagine a world which is a physical duplicate of actuality yet lacks mental properties. This is because, as we have seen, for all they know a priori, the physical duplicate may not be a causal duplicate. And it is only causal duplicates which are guaranteed, according to analytic functionalism, to be mental duplicates.

There is a reply, though, which I am sure the astute reader will already have considered: we can reformulate the conceivability argument in such a way as to consider *causal* zombies.

Analytic functionalism says that a causal duplicate is a mental duplicate. Now we consider a world which is a physical and causal duplicate of actuality and ask ourselves if we can conceive of it as lacking mental states: it's a zombie world.

If we can conceive of this, and if conceivability is a good guide to possibility, then there is a possible world just like our world physically *and causally* and yet lacks mental states. Thus, whatever it is about our world that gives it a mental nature must outstrip the combination of its physical and causal nature, and so dualism is true. It makes no difference what causality is, in the actual world, for we stipulate that whatever it is actually, the duplicate has that thing, and yet lacks mentality.

The objection is a good one: but to see why it's not dialectically apt we need to think about the relationship between the proposal of this chapter, and my original (Braddon-Mitchell 2003) proposal.

The objection does work against the 2003 proposal. For according to that proposal, the conditionality applied directly to the concepts of (at least phenomenal) mental states. The conditional concept of 'qualia' explained why someone who takes it to be the case that there are only physical states in the actual world should grant that they are sufficient for mentality, while accepting that if there are 'spooky' states in the actual world, they should regard those as necessary for phenomenal mental states, and thus accept that there could be a physical duplicate of the actual world which isn't a mental one. Simply adding that the physical duplicate of the actual world is also a causal one makes no difference to the reply.

That reply, though, required one to accept the conditional concept of qualia from 2003 (and the similar one in Hawthorne 2002). The current view is directed towards analytic functionalists (as perhaps it is right to think of Armstrong) who are happy with their analysis of mental states, but still think that somehow it is compelling that a physical duplicate of the actual world might fail to be a mental one.

But to the extent that they are happy with their analysis of mental states, the flat-footed response to the conceivability argument that seemed troubling in the case of zombies who are merely physical duplicates, should not seem so troubling in the case of physical and causal duplicates. They just need to deny that it is conceivable that there could be a causal duplicate of the actual world which fails to be a mental duplicate. For, by their lights, what it is to be a mental state is to play certain causal roles. It's a conceptual truth that anything that plays those roles is the relevant mental state, so, of course, it's inconsistent with that conceptual truth that there could be a world in which all the same things *happen*, in which the same internal states are related to the same external states causally and there be no mental states. If there is something which seems like 'conceiving' a causal duplicate which is not a mental duplicate, it's not ideal conceiving. It's just failing to comprehend the full causal nexus and how it relates to the true a priori story about the mental.

So, to recapitulate, it's a dialectical issue. My causal conditionalist view is for the benefit of someone who likes analytic functionalism and does not want to accept the conditional view of qualia. Such a person faces a puzzle explaining why there cannot be physical duplicates which are zombies. The present view explains that. Of course, it isn't in the business of explaining away why there can be zombies who are causal and physical duplicates. The analytic functionalist thinks that is a straightforward a priori contradiction, so is, of course, not ideally conceivable. If you don't think it's a straightforward a priori contradiction, then you aren't an analytic functionalist. Perhaps, then, you prefer the conditional qualia view, according to which analytic functionalism is true as a second-best deserver, only if the actual world lacks dualistic states.

10. Conclusion

Perhaps the Armstrong of *MTM* is an analytic functionalist. At very least, the causal theory therein is similar enough that it stands or falls with it when pitted against the conceivability argument. One solution is the conditional qualia view; but the Armstrong of *MTM* would have little truck with that view, which gives a kind of 'best deserver' status to dualism while accepting that in fact mental nature may be entirely causal-functional. But that view denies the straightforward conceptual discovery that the essence of mental states is exhausted by their causal roles. On that view, while this may be true, it's not straightforwardly a priori that it's true. So, how is the materialism of the causal theory of the mind to deal with the idea that perhaps physical duplicates could fail to be mental duplicates? I offer the view of this chapter as advice to anyone who wants to retain analytic functionalism or the causal theory in something close to its pure form, and still have a reply to the conceivability argument.

References

Alter, Torin (2007) 'On the Conditional Analysis of Phenomenal Concepts', *Philosophical Studies*, 134: 235–53.

Alter, Torin (forthcoming) *The Matter of Consciousness*, Oxford: Oxford University Press.

Armstrong, D. M. (1962) *Bodily Sensations*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1983) *What Is a Law of Nature?*, Cambridge: Cambridge University Press.

Boghossian, Paul and Christopher Peacocke, eds (2000) *New Essays on the* A Priori, Oxford: Oxford University Press.

Bourne, Craig (2002) 'When Am I? A Tense Time for Some Tense Theorists?', *Australasian Journal of Philosophy*, 80(3): 359–71.

Braddon-Mitchell, David (2003) 'Qualia and Analytical Conditionals', *Journal of Philosophy*, 100(3): 111–35.

Braddon-Mitchell, David (2004) 'How Do We Know It Is Now Now?' *Analysis*, 64(3): 199–203.

Braddon-Mitchell, David and Frank Jackson (1999) 'The Divide and Conquer Path to Analytical Functionalism', *Philosophical Topics*, 26(1–2): 71–88.

Campbell, Keith (1970) *Body and Mind*, New York: Anchor Books.

Chalmers, David J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press.

Forbes, Graeme A. (2016) 'The Growing Block's Past Problems', *Philosophical Studies*, 173(3): 699–709.

Forrest, Peter (2004) 'The Real but Dead Past: A Reply to Braddon-Mitchell', *Analysis*, 64(4): 358–62.

Gold, T. (1962) 'The Arrow of Time', *American Journal of Physics*, 30: 403–10.

Hawthorne, John (2002) 'Advice for Physicalists', *Philosophical Studies*, 109: 17–52.

Maudlin, Tim (2007) *The Metaphysics Within Physics*, Oxford: Oxford University Press.

Price, Huw (1997) *Time's Arrow and Archimedes' Point*, Oxford: Oxford University Press.

Ravenscroft, Ian and Stephen P. Stich (1994) 'What Is Folk Psychology?', *Cognition*, 50(1–3): 447–68.

Strawson, Galen (1994) *Mental Reality*, Cambridge, MA: MIT Press.

Tooley, Michael (1987) *Causation: A Realist Approach*, Oxford: Clarendon Press.

13

A Materialist Reconception of the Mind

Amie L. Thomasson

David Armstrong, like so many of us, worried about the lack of progress in philosophy. 'The notorious fact about philosophy is its failure to make the sort of progress made in science . . . The goal of philosophy is knowledge but, unlike science . . . philosophy does not seem to attain knowledge. This has been the great scandal of modern philosophy . . .' (Armstrong 1973, 21). While scientific activity has 'as its end the compiling of a book or encyclopedia' of settled knowledge, Armstrong lamented that by contrast, '[p]hilosophers write books, but philosophy has no book' (Armstrong 1973, 21).

How should we respond to this 'scandal'? One promising option is to insist that metaphysics should not be seen as a rival to the sciences, or as a quasi-science engaged in trying to discover truths, to write a 'book of the world'—as Armstrong despaired of it being able to do.[1] Instead, we may see metaphysics as centrally engaged in both descriptive and normative conceptual work—work one might call 'conceptual research and development'.[2] That is, metaphysics often (implicitly or explicitly) engages not only in *descriptive* work concerning how our actual conceptual scheme works, but also *normative* work in assessing what concepts or terms we *should* use, and how we *should* use them. And while that may not be all it historically has aspired to do, it is something it legitimately can do—and which may crucially impact how we talk and live and how we understand our world.

Rethinking metaphysical work in these terms is appealing in many ways. It avoids the threat of a rivalry with science—a rivalry metaphysics seems bound to lose. It also makes the inevitable pluralism and divergencies we see in philosophical theories not an embarrassment that shows failure to converge on the truth, but a richness of proposals from which to choose for various purposes.

Huw Price (2011, Chapter 9) gives us another good reason for starting with the conceptual when we investigate what we would normally think of as metaphysical issues. As Price argues, those who aim to have a clear naturalist methodology should take the time to investigate how different areas of discourse are actually

---

[1] For these arguments, see Thomasson (2017; 2018; forthcoming), and the first of these Anderson lectures. The phrase 'book of the world' is from Sider (2011).

[2] The phrase 'conceptual research and development' has been attributed to C. B. Martin, though I have not been able to find confirmation of that in print.

Amie L. Thomasson, *A Materialist Reconception of the Mind* In: *Armstrong's Materialist Theory of Mind.*
Edited by: Peter R. Anstey and David Braddon-Mitchell, Oxford University Press. © Amie L. Thomasson 2022.
DOI: 10.1093/oso/9780192843722.003.0014

used, and what function(s) they serve, instead of assuming that all areas of discourse serve to track or describe certain features of the world (2011, 20–3). If our conceptual research turns up good reason for thinking that a range of vocabulary, say, the moral or the modal, does not have a describing or tracking function, then we have good reason to not start with the metaphysical question: 'What are moral or modal properties?', 'What is their relation to physical properties?' and the like—as if we could just investigate things we are tracking and learn their nature. Instead, we may get further if we first look at how the discourse functions, what it does for us, what rules it follows, and how it may eventually enable us to speak of the relevant facts or properties. Starting with the metaphysical questions presupposes a certain uniform answer to the linguistic functional question—one that may in some cases lead us astray.

The problems of philosophy of mind arose from naturalistic inclinations—aiming to get an account of mind consistent with modern science. Perhaps (following Price) they can be better handled by carrying over that naturalistic inclination to our study of talk about the mental, rather than assuming all discourse serves the same descriptive function.

What does philosophy of mind come to look like, if we look at it in these terms? What does Armstrong's contribution look like, if we look at it in these terms? Are there advantages to be gained by reorienting the work of philosophy of mind in this way?

If we look at it in this way, the first question to ask in the philosophy of mind is not the metaphysical question 'What are mental states?' or even related questions, such as 'Are mental states physical states?' or 'Are there mental properties that are not reducible to physical properties?' or 'Can mental states be causally efficacious?'

Instead, we should begin with questions relevant to conceptual research and development. What is mental discourse for? What does it do for us? What can we do, by having mental terms in our vocabulary, that we could not do—or couldn't do so easily—without them? How do mental terms work? Does using these terms in standard ways commit us to the existence of something immaterial, non-physical, something at odds with modern science or physicalism? If so, do we need to alter these concepts (how?), or replace them? If not, how *does* it work?

In fact, I think we should alter the statement of the project a little more—for perhaps we shouldn't assume at the outset that all mental terms are alike, that all have the same function or work in the same way. Terms we classify as 'mentalistic' may have little more in common than animals we classify as sea creatures, and we may be equally led astray if we presuppose a unity here and try to form generalizations in these terms. Perhaps we should begin by looking separately at our terms for beliefs, desires, imaginings, sensations, after-images, emotions, and so on. I will come back to this point later.

1. Re-conceptualizing Armstrong's Project

What I want to explore here is a way of re-reading Armstrong's project in *A Materialist Theory of the Mind* (hereafter *MTM*) in the above terms. I do not mean to suggest that Armstrong explicitly thought of what he was doing in these terms. On the contrary, he seemed to be always object-oriented in his thinking, and (despite his many merits) to be constitutionally blind to possibilities of functional pluralism in language.[3] Nonetheless, re-conceiving what he was up to in this way can make sense of some otherwise puzzling features of the text, and can lead us to reassess the achievements of the work.

It's worth beginning by looking at the fascinating Preface added to the paperback edition of *MTM* (Armstrong 1993). On the one hand, there we can see Armstrong himself dither about what he was up to in the book. As he reminds the reader there, he put forward the theory as a 'theory about concepts, as a set of logical analyses' (Armstrong 1993, xv), noting apologetically that '[t]his was probably the result of the Oxford part of my time as a philosophy student (1952–4). I would be much less dogmatic now' (Armstrong 1993, xv). 'Perhaps it is just a theory about the mental instead of an analysis of our concepts of the mental . . . [it] may be, partially at any rate, an empirical theory. I do not find the matter as important as I once did' (Armstrong 1993, xv).

Why does Armstrong come to dither about the nature of his project? He is often impatient with what he considers the outmoded way of thinking of philosophy as (merely) conceptual analysis—he mocks Ryle and Wittgenstein on this score (*MTM*, 55), and alludes to his own change of heart (Armstrong 1993, xv). Conceptual analysis of course faces familiar difficulties.[4] And even Armstrong himself comes to doubt whether some of his conceptual analyses succeed, writing in the new Preface: 'Many have wondered, I among them, whether the theory really captures the essence of mentality' (Armstrong 1993, xviii). His central move in *A Materialist Theory of the Mind*, of course, is to analyse mental concepts as *the concept of being an inner state apt for producing certain kinds of behaviour* (*MTM*, 82)*—and, in some cases, apt for being produced by certain ranges of stimuli* (*MTM*, 116). While Armstrong attempts to analyse all mental state concepts in this way, however, he realizes that this is a hard road to take for certain mental state concepts, including those of perceptions, sensations, and images (*MTM*, 123).

---

[3] This blindness shows up, for example, in his excoriation of Ryle for failing to say what the truthmakers for dispositional statements are (Armstrong 1993, xvi; 2004, 2–3)—letting dispositions 'hang on the air'. But this criticism shows a failure to appreciate Ryle's central move of suggesting that mental talk has a non-descriptive function (see Ryle 1949, 116–49). It also shows up in his apparent misconstrual of Feyerabend's eliminativism (*MTM*, 78), and in his insistence that it is proper to ask, for any truth, what the truthmaker is (Armstrong 2004).

[4] Famously including those coming from Quine's critique of the notion of analyticity (1953/2001) and from Kripke's (1980) and Putnam's (1975) work on theories of direct reference.

But the choice Armstrong looks at here, of whether his theory is better thought of as conceptual analysis or as an empirical theory, may (to use a phrase of Ryle's (1970, 10–11)) have 'too few "ors" in it'. There is another option for how we can understand the basic drive behind his project—one which Armstrong himself did not explicitly articulate. We can understand it as a project in conceptual ethics.

What is that project, why should we think of Armstrong's work in that way, and what changes if we think of Armstrong's work—and perhaps other work in philosophy of mind—in these terms?

The fundamental problem we can then see as driving the work of Armstrong and many others of his era in philosophy of mind (and the generation before him) is this: How can and should we conceive of the mental, in a way that will be coherent with modern scientific knowledge? What can and should we do to understand our talk of the mental in a way consistent with a modern scientific world-view?

So understood, this is a problem neither of conceptual analysis (thought of as: What is the most accurate analysis of how our actual mental concepts work?), nor of empirical research. It is rather a problem for conceptual ethics and (perhaps) conceptual (re-)engineering. That is: what should we do with our mental concepts? It is, of course, a problem that arises for *empirical reasons*. Empirically (as Armstrong would have been the first to emphasize), we have more and more reason to accept a scientific view of the world, to think that there are no non-material substances, to think that neuroscience and psychology may be able to give good accounts of our behaviour. Addressing the problem well also relies on conceptual analysis—are our extant mental concepts already consistent with a materialist world-view—or not? So, there is an element of truth to thinking that both conceptual analysis and empirical information play important roles. But, strictly speaking, the problem is neither an empirical one, nor a descriptive conceptual one. At bottom, the problem is (given advances in the sciences and all we know about the brain and behaviour now): What *should* we do with our mental concepts?

So understood, the project is continuous with what one might think of as the overarching project driving much of philosophy since the modern period: What shall we do with our conceptual scheme, in light of the advances of modern science, and the retreat of religion? What should we do with our concepts like that of *person*, *freedom*, *causation*, *right and wrong*, *responsibility*, *object*, *solidity*, *space*, *time*... ? Can they be understood, in their ordinary sense, in some way that doesn't appeal to souls, immaterial selves, divine action or teleology, or a unified plenum of stuff? Or do we need to amend them, reinterpret them, reconfigure them, replace them with scientific surrogates, or eliminate them?

Understanding the mind is just one part of this larger project of working out what is to be done with our conceptual scheme as a scientific world-view continues to develop. As Armstrong puts it:

> [We have] a very simple reason for giving special weight to scientific considerations. Historically, scientific investigation has proved to be the only way that a *consensus* of opinion about disputed matters has ever been achieved among those who have given the matters serious and intelligent attention. Only science has settled disputed questions. This is why the seventeenth century is a landmark in the intellectual history of the human race. (*MTM*, 52)

This way of thinking of the project is totally coherent with the way Armstrong characterizes his project in the original edition of *A Materialist Theory of the Mind.* The driving question, with which he introduces the book, is: 'What is a man?... Is man nothing but his material body? *Can we give a complete account of man in purely physical terms?*' (*MTM*, 1, emphasis added).

Why does this question arise now?

> More and more psychologists and neurophysiologists explicitly or implicitly accept the view that, so far as mental processes are concerned, there is no need to postulate anything but purely physical processes in man's central nervous system. If we take the word 'mind' to mean 'that in which mental processes occur' or 'that which has mental states', then we can put this view briefly and not too misleadingly as: the mind is nothing but the brain. If scientific progress sustains this view, it seems that man is nothing but a material object having none but physical properties. (*MTM*, 1)

The aim of the book, as Armstrong presents it, is to show that there are 'no good philosophical reasons for denying that man is nothing but a material object' (*MTM*, 2). And he considers this to be important since it may remove obstacles to scientific progress. The view, if it's true, (Armstrong says) is one for which science (in particular, neuroscience) must provide the proof—he aims only to remove the 'harassing action' of philosophers (*MTM*, 2). We can see this goal also in a central criticism he raises of the Cartesian view. Cartesian dualism, he suggests, is a coherent option for understanding our mental concepts, but 'highly implausible from a scientific point of view' (*MTM*, 30), given that it seems unable to account for the gradual development of mental processes, and requires a scientifically implausible account of causation between the mental and physical, as the existence of special laws connecting physical to mental entities 'fits in very ill with the rest of the structure of science' (*MTM*, 50).

The project, if we think of it in these terms, is at bottom a matter of giving a way of thinking about the mind—a way of handling our mental concepts—that will be consistent with the view that we are simply material beings in a material world, with physical properties. Interestingly, J. J. C. Smart—who Armstrong says converted him to the view that 'mental states are nothing but physical states of the brain' (*MTM*, xi)—credits Rudolf Carnap (1932) with one of the earliest versions of an identity theory, though he notes, 'Carnap did regard the identity as a linguistic recommendation rather than as asserting a question of fact' (Smart 2000/2017, Section 1; see also Carnap 1963, 886). This should come as no surprise given Carnap's well-known standing as the grandfather of contemporary work in conceptual engineering. So, in short, the very 'identity theory' that Armstrong develops began life as a *linguistic recommendation* in the work of Carnap (1932) and other logical positivists (such as Schlick (1935)), who aimed to ensure psychology's place as a science by enabling it to employ an intersubjective terminology. Indeed, as Sean Crawford (2013) nicely details, there were originally two different versions of a 'physicalist' theory. The first, which Crawford labels (capital 'P') 'Physicalism', arose with the logical positivists (prominently including Carnap) in the 1930s, who took the mind-body problem to be a metaphysical pseudo-problem, and recommended a choice of language that would 'translate' psychological terms into (extensionally equivalent, but perhaps not meaning-equivalent) physical terms, as a way of securing psychology's intersubjectivity and its place as a science. The second (small 'p') 'physicalism', made prominent in the 1950s with work by Place, Feigl, and Smart (but apparently originated by a pre-positivist Schlick (1925/1974)), and followed up later by Armstrong, reconfigured physicalism as an ontological thesis: not treating the mind-body problem as a pseudo-problem but rather as a genuine ontological problem, and defending a solution to it.

But even in this new context, the alternative of thinking of the project in terms of *what conceptual scheme we should choose* remained alive. Smart himself seems to have been a transitional (or equivocal) figure. For he notes, in a direct echo of the positivist program, that one of the central considerations that *inclines* philosophers like him to accept the 'brain process thesis' is the unification of science (1967, 84–5) (the other consideration he cites is Occam's razor).[5] Moreover, after his 'topic-neutral' analysis of sensation reports was criticized (by M. C. Bradley) as circular, Smart retreated to suggest instead adopting Paul Feyerabend's (1963) view that (even if physicalism is not a common-sense view) the 'conceptual scheme' of common sense 'should not be treated as sacrosanct' (Smart 1967, 91). Instead of treating physicalism as an ontological hypothesis, he suggested, we can 'go in the direction of simply *replacing* dualistic conceptions with physicalist ones' (Smart 1967, 91), changing our conceptual scheme.

To be clear, again I don't think that Armstrong explicitly thought of his project as one in conceptual ethics—I don't think he saw the possibility of this option, as he tended to think exclusively in metaphysical terms. But it was an option that *was* prominent in the earlier history of physicalist theories in the philosophy of mind that Armstrong builds on. It was also an option that remained on the table in work in his circle in the 1960s, as the work by Smart and Feyerabend shows.

---

[5] As Crawford notes, following Leopold Stubenberg, the Australian Place-Smart (and later Armstrong) versions of an ontological identity theory were again importantly different from the Austrian (early) Schlick-Feigl theory, since the former treated the mentalistic vocabulary as the problematic side, while the latter considered the physicalistic concepts as those requiring 'reinterpretation and reconstruction' (Crawford 2013, 647).

If we think of Armstrong's project as implicitly concerned with reconceptualizing the mental in a way that makes it coherent with a modern scientific (and neuro-scientific) view, we can better capture what he was up to, in a way that fits with his description of his project and explains his dithering over its nature. If this is what Armstrong was implicitly up to, it makes sense of why Armstrong wasn't quite satisfied with thinking of his own work either as conceptual analysis or as empirical work. In any case, whether or not Armstrong explicitly thought of his project in this way, it is clearly a worthwhile project to which he may have made crucial contributions.

But there is one aspect that remains mysterious, if we think of Armstrong's project in these terms: Why is he then so hostile to eliminativism? For eliminativism, one might think, gives us one way to acquire a conceptual system coherent with modern science: by making a move suggested by Feyerabend and at least partially endorsed by Smart above—eliminating our mental concepts (or at least the positive atomic use of some sticky ones, such as sensation reports) altogether, and replacing them with physical concepts (say, those of neuroscience). In short, if the goal were just to gain a conceptual system coherent with modern science, eliminativism would seem to provide a way to do this. Armstrong even describes Feyerabend's eliminativist view in these terms:

> one physicalist, Paul Feyerabend, . . . has suggested that the materialist ought simply to recognize that his world-view does not allow statements that assert or imply the existence of minds. A true physicalism will simply talk about the operation of the central nervous system, and will write off talk about the mind as an intellectual loss. (*MTM*, 78)

If the concepts of mental states are simply the concepts of being an inner state apt for producing a certain range of behaviour (and/or being produced by a certain range of stimuli), and neuro-scientific concepts ultimately turn out to provide a better explanation of the connections between stimulus and behaviour, why not be open to replacing them?

Yet Armstrong writes of this: 'I think that if the situation is as desperate as this it is desperate indeed' (*MTM*, 78). In the second edition Preface to *A Materialist Theory of the Mind*, he again expresses hostility to eliminativism, writing:

> One Materialist theory that I have never been drawn to is the Eliminativist account of the mental . . . If I were to become convinced that there is an incompatibility between a materialist or physicalist view of the world and the existence of the mental, then I would reluctantly turn Dualist. Materialism is a theory, even

> if, as I think, a good theory. The existence of mental things—pains, beliefs and so on—seems to me to be part of bedrock, Moorean, common sense. Its epistemic warrant is far better than that of Materialism. (Armstrong 1993, xix)

There are two ways to understand eliminativist views—about the mental or anything else. One is as the simple *descriptive* claim, made *using* standard mental state terms, that there are no mental states. It is this that seems to be a violation of 'bedrock, Moorean, common sense' (*MTM*, xix). This seems to be the way Armstrong is understanding it—an interpretation that again bears witness to Armstrong's tendency to think always in metaphysical terms (in what Carnap would have called 'the (misleading) material mode').

The other way of understanding certain forms of eliminativism is as a (either implicit or explicit) *recommendation* or *proposal* that we replace a range of discourse (in this case, discourse involving our traditional mental concepts) with a different range of discourse (in this case, perhaps, neuroscientific discourse)—much as we might replace our talk of warmth with more precise talk of temperature (see Carnap 1950, 12–13). The Feyerabend whom Armstrong is quoting is taking the latter tack—arguing explicitly that it's a mistake to treat statements that identify a kind of mental process with a kind of physical process as stating empirical hypotheses. Instead, he argues, the 'proper procedure' for monists is to develop their theory 'without any recourse to existent terminology' (Feyerabend 1963, 295), using such identity statements, if at all, for 'redefining "mental process"', not for expressing an empirical hypothesis (Feyerabend 1963, 296).

Armstrong is in a perfectly good position to resist the first sort of eliminativism, but on much weaker ground in resisting the latter. (Perhaps he rejects it so vehemently because he is only understanding it in the first way—again, perhaps, part of his habit of thinking metaphysically and treating all discourse as describing.) Armstrong is on weak ground in resisting eliminativism, taken as a *proposal*, since, on his view, 'The concept of a mental state is the concept of that, whatever it may turn out to be, which is brought about in a man by certain stimuli and which in turn brings about certain responses. What it is in its own nature is something for science to discover' (*MTM*, 79). If the concepts of mental states are just concepts of states (of whatever kind) that mediate certain kinds of stimuli and certain kinds of behaviour, and the goal is simply prediction and explanation, then it's unclear why we shouldn't be open to shifting to a more precise and explanatory scientific vocabulary—to that, say, of neuroscience, given that 'Modern science declares that this mediator between stimulus and response is in fact the central nervous system' (*MTM*, 79). Of course, there might remain avenues of defence, say, if mentalistic explanation provided a sort of coarse-grained explanation that remained more handy in everyday contexts and that was unavailable in neuroscientific terms. But that is a case that would have to be made, and to which Armstrong does not even allude.

I think there is actually a good response available to the question of why one should retain the mental concepts and resist the eliminativist impulse, even when we take it in the form of a proposal. However, the response is one that Armstrong is not particularly well positioned to appreciate. The response is that, at least for some kinds of mental terms, it may not be an apt or complete account of their function to say that they serve to name inner states that can predict (and explain) behaviour. This harkens back to Strawson's (1963) complaints against Carnap's attempts to engage in philosophical 'clarification' by replacing imprecise everyday concepts with scientific ones:

> The scientific uses of language . . . are extremely highly specialized uses. Language has many other employments. We use it in pleading in the law courts; in appraising people's characters and actions; in criticizing works of art; in recounting our states of mind . . . it seems in general evident that the concepts used in non-scientific kinds of discourse could not literally be *replaced* by scientific concepts serving just the same purposes; that the language of science could not in this way *supplant* the language of the drawing-room, the kitchen, the law courts and the novel. (Strawson 1963, 505)

Along these lines, Kristin Andrews brings empirical evidence that belief attributions may fulfil other pragmatic goals: 'we explain behavior to impress other people, to condemn other people, and even to reduce the discomfort associated with having seemingly inconsistent beliefs about a person' (Andrews 2012, 116).

Interestingly, behavioural explanations appeal explicitly to the actor's beliefs and desires most often in two sorts of cases: first, 'when they are motivated to portray the behavior in a positive light', and, second, when they are explaining their *own* behaviour (Andrews 2012, 111). Assuming that we are regularly motivated to portray ourselves in a positive light, there seems to be a unity behind these cases: attributions of propositional attitudes are used to explain behaviour in order to portray it positively—to *justify* behaviour.

Andrews suggests that 'a better account of the function of mindreading is the development of sophisticated moral abilities such as the justification of behaviour, which helps to explain both how a theory of mind facilitates group living and how it facilitates the development of technological advances' (Andrews 2012, 218). The function of evolving a theory of mind, she argues, is offering *reasons,* not predictions, particularly in anomalous cases where some norms are violated. On Andrews' view, belief attributions presuppose a background system of norms (Andrews 2012, 222), and are made to explain norm violations (Andrews 2012, 224).

The ability to justify our own actions and those of others might itself be useful (Andrews argues) in that justifications of violations of social norms enable other group-members to understand and accept innovations (including technological innovations). One might also more broadly appeal to a reduction of conflict among group members—for if someone's action, which at first looked simply cheeky, offensive, or destructive, comes to be understood as reasonable, given the agent's beliefs and desires, rancor is reduced, and people are better able to live together. One can see this also in the typical difficulties faced by those on the autism spectrum, who often have trouble maintaining social relations and avoiding conflict—difficulties that go along with difficulties in making standard attributions of beliefs, desires, etc. to their neurotypical peers. If our belief attributions serve to justify actions in ways that enable us, as social creatures, to better live together, then we would do well not to give them up, even if we come to acquire neurological concepts that allow for more precise behavioural predictions.

2. Reconceptualizing Philosophy of Mind

This brings us back to the idea I mentioned at the beginning—that if we are to address questions in philosophy of mind as questions in conceptual ethics, we should begin by asking what such terms *do for* us. And in asking these questions we should not lose Ryle's lesson (lost with Armstrong) that we should not simply assume that all terms serve to refer to and describe (inner or outer) things.

I proposed above taking a fundamental question of philosophy of mind to be 'What should we do with our mental concepts (given what we now know about the workings of the brain, the causes of behavior, and so on)?'. But before we determine what we should *do with* our mental concepts, we should investigate what they do *for* us.

It is important to note that, in asking about the functions of terms, we needn't presuppose that the function is an overall *beneficial* one—or that the terms should be preserved rather than eliminated. Traditional race terms, for example, served to give pseudo-scientific legitimacy to discriminatory treatment, and terms like 'Hochdeutsch' might serve to reinforce regional and class biases and the dominance of one linguistic group at the expense of others. Nor need we suppose that terms always succeed in fulfilling their functions: 'phlogiston' was a term introduced to serve in tracking a chemical kind released in combustion, in a way that could serve well in our chemical theories—but it turned out to fail in that function, and is best eliminated from our chemical vocabulary. We can thus ask about the functions of a range of vocabulary without pre-judging whether that vocabulary should be eliminated. Terms might be worth eliminating, for example, if they serve a function we (now) reject or see as harmful, or if they fail to properly serve a legitimate or desirable function, or if their functions could be performed better by a set of replacement terms.[6]

---

[6] See Thomasson (forthcoming) for further discussion.

How might work in philosophy of mind might look different if we re-conceptualize it in this way? As I mentioned above, I think that we shouldn't presuppose that all mental terms are alike, or function in the same way. So I will just treat a single example here—a case where Armstrong's attempt to give a reconceptualization proved to be problematic, and yet where an alternative functional story might prove helpful. That case is talk of appearances or seemings.

As Armstrong notes, once we realize that one may have perceptions as-if there were a red ball, even when there is no ball there, 'it is very tempting to say that what is involved is some relationship between my mind and a non-physical red item: a sense impression or sense-datum' (*MTM*, 217). But if that is the right analysis, then he cannot preserve his account of perceptions as states of a person apt for bringing about behaviour, and/or for being brought about by certain stimuli. For one may be 'hip' to the illusion involved and not be apt to act on it, and it need not be brought about by the usual stimuli.[7] Armstrong admits that he has difficulties accounting for perceptual experience in a way that is compatible with his causal analysis. He describes a number of difficulties with positing 'sensory items' that would be the referents of these terms (*MTM*, 217–21), and proposes instead that these are cases in which we have events that 'would be the acquiring of belief, but for the existence of other, contrary, beliefs that we already hold' (*MTM*, 223)—thus remaining consistent with his causal theory.

Now this account is one that might, to many, seem to leave something out in our understanding of the way things appear or seem to us: it leaves out anything like what we would call 'phenomenal content'. This is one of those cases where, as William Kneale put it in his review, Armstrong is 'more anxious to prepare the way for his identity thesis than to report facts as he finds them' (Kneale 1969, 299)—and Armstrong seems to know it.

But suppose we turn from the thought that appearance or seeming talk must be describing certain immaterial 'sensory items', to look back at how such talk is *used*, and how it may have been introduced into our vocabulary. Smart (unlike Armstrong after him) is very much attuned to the option of treating appearance talk as *doing something other than describing.* In the initial pages of 'Sensations and Brain Processes', he explains and makes clear that he is 'very receptive to' an expressivist view of sensation statements (Smart 1959, 144). That is, he is sympathetic to the view that in talking of having a roundish yellowy orange after-image, we might not be 'reporting' anything, but rather 'expressing some sort of *temptation,* the temptation to say that there *is* a roundish yellowy orange patch on the wall (though I may know that there is not such a patch on the wall)' (Smart 1959, 141).[8]

---

[7] For the terminology of 'hip' hallucinations and helpful discussion, see Smith (1983).

[8] Smart also there says that he finds 'congenial' what he takes to be Wittgenstein's expressivist view of pain talk—that when I speak of having a pain I am really 'doing a sophisticated sort of wince' (Smart 1959, 141).

And this expressivist option, as Smart notes, would enable us to retain a respectably scientific view of the world, resisting the idea that to say '"I have a yellowish orange after-image" is to report something irreducibly psychical' (Smart 1959, 142). In the end, of course, Smart suggests that he does 'not feel that [the expressivist account of sensation statements] will quite do the trick', since he feels that there is *some* sort of genuine reporting involved—and that there aren't the presumed barriers to thinking that sensation reports are reports of brain-processes (Smart 1959, 144). But he was certainly alive to the possibility of a non-descriptive account, and saw clearly that a successful non-descriptive account would fulfil his deepest goals of giving a scientifically respectable account of mental discourse.

There are, in any case, other non-descriptive accounts of appearance talk to be considered. Wilfrid Sellars (1956/2000, 50) developed the interesting suggestion that talk of *seemings* or *appearances* shouldn't be taken as aiming to describe inner states, but rather as enabling the speaker to withhold world-regarding commitment, a way to hedge our full endorsement. So when John, the tie salesman, shifts from saying 'the tie is green' to 'the tie appears green' or 'the tie looks green to me', what he is doing is hedging, withholding commitment to the worldly claim about the tie, making himself less liable to criticism in the event that his interlocutor takes the tie outside and is inclined to claim that it is blue. Plainly, this kind of commitment withholding, useful in our linguistic scorekeeping practices, in what we hold speakers responsible for, can also help reduce conflict in the social world. This also captures the idea that in shifting to appearance talk, we reduce our liability to error. For when we are less committed, we are less open to mistake. And it fits well with the idea of transparency, for we think of our introspective reports not as reports of additional quasi-perceptual states in which we perceive the states of our own minds, but rather as ways of withholding commitment from the first-order worldly reports.

Sellars' suggestion, as I have argued elsewhere (Thomasson 2005), was inspired by Edmund Husserl's broader use of a method of phenomenological reduction to enable us to talk not about the world, but about the ways in which we come to be presented with the world—enabling us to acquire knowledge in phenomenology. It is worth noting that if Husserl's method of bracketing is possible at all (a kind of suspension of judgment about anything to do with the world represented), then there are perceptual states which (bracketed) have no tendency to affect our acquiring beliefs or acting in any way—they are 'disconnected'. In Husserl's hands, we can get to knowledge of *phenomena*, ways in which the world is presented to us, by 'bracketing' the assumption that our judgments are true, and that our experiences are veridical. This bracketing enables us to introduce phenomenological terms—'seems' talk and 'appearance' talk.

Ultimately, as I have argued elsewhere (Thomasson 2005), we can become entitled to move from the reductive transformation that takes us from 'the tie is green' to 'the tie seems green' to a hypostatizing transformation that enables us to refer to appearances—in this case, to 'there is an appearance as-of a green tie'. Engaging in bracketing like this, which enables us to introduce terms for how things *seem*, or more broadly, for *appearances*, does not require 'self-scanning', and does not involve reporting on observed 'sensory items'. Instead, in Husserl's hands, it serves a transcendental function, of enabling us to acquire knowledge of how the world is presented to us (or constituted by us) in experience. So understood, we retain the advantages of respecting transparency (for what we have is a cognitive transformation from a first-order, world-oriented experience, not a new experience of a new sort of thing), and of respecting the idea that there is some protection from error—for the transformed statement 'there appears to be a green tie' may be true even if the original world-directed experience was not veridical (Thomasson 2005, 134).

Suppose that something like the Sellarsian/Husserlian story about appearance talk is on the right track. If so, then appearance talk does not enter language as an attempt to describe non-physical states or 'sensory items' (in which case we must either reject materialism or reject or reconstrue the language). Instead, it arises as a way of withholding world-regarding commitment, which in turn enables us to come (through simple cognitive transformations) to acquire knowledge of our ways of representing a world. The ability to conceptualize and speak of appearances may also be useful for us in daily life—for only if we have a way of conceptualizing or speaking of appearances can we distinguish appearance from reality and keep track of how things appear to others (regardless of how they are in the world—enabling us to pass 'false belief' tests).

So, suppose we think of 'appearance' talk as introduced with rules roughly like these, to serve functions like those considered above, and come back to the general question of what we should do with our mental concepts, given modern scientific knowledge. Is the kind of language use involved in appearance talk in tension with a scientific world-view? I don't see why it would be—it certainly doesn't pretend to be 'tracking' any 'immaterial states' or 'sensory items'. It also doesn't require any mysterious account of how we can come to know them. Instead, it simply shows how we can introduce a new kind of vocabulary with a commitment-withholding function, and transform that into noun form, enabling us to talk of, and (non-mysteriously) acquire knowledge of, our ways of representing or 'constituting' a world. In this way, it runs somewhat parallel to the way expressivist accounts of moral talk, or normativist accounts of modal talk, account for the relevant forms of discourse without 'positing' moral or modal 'items' to explain it.

Now this is just an example—I don't aim here to fully develop or defend this approach to 'looks', 'seemings' and 'appearance' talk, or to speculate on to what extent the approach may help us understand other kinds of mentalistic vocabulary.

But I think it is clear is that we do better if we begin work in philosophy of mind a step back from its usual starting place: asking not what mental states (or beliefs, or sensations, etc.) *are* (and how they relate to the physical), but asking whether and how we can understand mental discourse of various sorts in ways that fit well with a scientific world-view, and if not, what we should do about it. Seen in this way, the central problem in philosophy of mind is a problem in conceptual ethics.

3. Reassessing *A Materialist Theory of the Mind*

I suggested above that we try reconceiving of Armstrong's project as one in conceptual ethics: How can and should we conceive of the mental, in a way that will be coherent with modern scientific knowledge? What can and should we do to understand our talk of the mental in a way consistent with a modern scientific world-view?

Armstrong, I think, began by seeing the problem in roughly that way. But he tended to think that investigations into the mind must be *either* a matter of conceptual analysis *or* empirical investigation—a thought that had too few 'ors' in it. Appropriately enough, he found both options unsatisfying, and unable to quite capture what he was up to.

But if we reconceive of his project as determining how we *should* conceive of the mental, we can re-evaluate both the achievements and shortcomings of his work in *A Materialist Theory of the Mind.* First, he is right to insist on this project, as an essential one for philosophy—and one that becomes ever more urgent, the more we discover about the workings of the brain and the causal basis for our experience and behaviour. We can also see his analyses of particular mental concepts as whatever states are 'apt for producing certain behaviour' (and in some cases: apt to be caused by certain stimuli) as enabling us to have ways of speaking of the mental without commitment to the existence of anything immaterial, anything at odds with a materialist metaphysics. Moreover, and usefully from Armstrong's perspective, to the extent that we can see his analyses as conceptual *proposals,* they won't be open to criticism on the grounds of failing to capture how our actual mental concepts work.

On the other hand, once we notice the possibility of functional pluralism, some of Armstrong's limitations also come to light, since he seemed blind to this possibility. Perhaps that is why he didn't fully appreciate the difference between analysing our concepts and proposing reforms of them. In addressing talk of the mental, Armstrong similarly seemed constrained by a descriptivist assumption, apparently assuming that if a materialist world-view precludes taking mental state talk to be describing immaterial inner states, it must be describing physical inner-states—states of the central nervous system. As a result, he wasn't positioned to notice the difference between discourse that would *describe* causes versus discourse that would *license* inferences, *justify* actions, or *withhold* commitment. Not noticing these options limited him unnecessarily, providing him with too few 'ors' in the choices he saw as available to solve the problems in philosophy of mind.

Once we explicitly acknowledge the possibilities of functional pluralism in our discourse, we can see more options available. We can, for example, see new paths available for resisting eliminativism—even where it is presented as a proposal for reforming our language, rather than as an ontological 'discovery'. For there may be functions mental discourse serves other than aiming to 'posit' or 'track' entities to explain behaviour—functions we might lose by dropping traditional mental terms in favour of neurological explainers. Moreover, we may see new options for understanding certain of our mental concepts in ways that fulfil our goals in conceptual ethics, without having to do violence to our actual concepts. For example, as I mentioned above, Armstrong faced notorious difficulties with his analysis of non-veridical perceptions, or what we might call 'mere appearances'. Keen to avoid the idea that appearance talk aims to *describe* inner 'sensory items', he proposed that these are simply cases in which we acquire 'potential beliefs', overridden by other, stronger beliefs (*MTM*, 223). Yet even he was aware of the shortcomings of this, taken as conceptual analysis, admitting, 'We can quite well imagine the occurrence of perceptions that involve no acquiring of belief at all, even although contrary beliefs about the world are quite absent' (*MTM*, 223). But (as I have suggested above), opening up our range of options to consider other functions and inferential roles for appearance talk enables us to see how appearance talk could enter language and be used successfully, without either thinking of it as aiming to describe 'inner sensory items', or as analysable in terms of talk of potential beliefs.

In short, once we give up the descriptivist assumption, we can see more 'ors' available for characterizing the project, and more 'ors' in the options for answers. For we can treat the central problem as *neither* one of simple conceptual analysis, *nor* as one for empirical investigation, but rather as a problem in conceptual ethics: *What should we do with our mental concepts?* This requires first asking what various kinds of mentalistic discourse do for us, and whether they need reform to fit with our scientific knowledge. If we are attuned to the possibility of functional pluralism, we may see other options than just thinking of mentalistic discourse as describing immaterial or physical states—options that may do less violence to our mental concepts as we know them and use them, and that may relieve us of the need for reform or elimination. Nonetheless, if conflicts do arise that require reform, addressing the problems in philosophy of mind as problems in conceptual ethics also opens new avenues for solutions. For we needn't be constrained to just analysing our old concepts, nor need we present our philosophical work as if it reports new empirical discoveries. Instead, we may justify our proposals by explicitly showing the benefits to be gained by changing our conceptual system.[9]

References

Andrews, Kristin (2012) *Do Apes Read Minds?: Toward a New Folk Psychology*, Cambridge, MA: MIT Press.

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1973) 'Continuity and Change in Philosophy', *Quadrant*, 17: 19–23.

Armstrong, D. M. (1993) 'Preface', in *A Materialist Theory of the Mind*, 2nd edition, London: Routledge, xiii–xxiii.

Armstrong, D. M. (2004) *Truth and Truthmakers*, Cambridge: Cambridge University Press.

Carnap, Rudolf (1932) 'Psychologie in physikalischer Sprache', *Erkenntnis*, 3: 107–42. English translation in *Logical Positivism*, edited by A. J. Ayer, Glencoe, IL: Free Press, 1959.

Carnap, Rudolf (1950) 'On Explication', in *Logical Foundations of Probability*, Chicago: University of Chicago Press, 1–18.

Carnap, Rudolf (1963) 'Herbert Feigl on Physicalism', in *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schilpp, La Salle, IL: Open Court, 882–6.

Crawford, Sean (2013) 'The Myth of Logical Behaviourism and the Origins of the Identity Theory', in *The Oxford Handbook of the History of Analytic Philosophy*, edited by Michael Beaney, Oxford: Oxford University Press, 621–5.

Feyerabend, Paul (1963) 'Comment: Mental Events and the Brain', *Journal of Philosophy*, 60(11): 295–6.

Kneale, William (1969) 'Review of David Armstrong, *A Materialist Theory of the Mind*', *Mind*, New Series, 78(310): 292–301.

Kripke, Saul A. (1980) *Naming and Necessity*, Oxford: Blackwell.

Price, Huw (2011) *Naturalism without Mirrors*, Oxford: Oxford University Press.

Putnam, Hilary (1975) 'The Meaning of "Meaning"', in *Language, Mind and Knowledge*, vol. 7, *Minnesota Studies in the Philosophy of Science*, 131–93.

Quine, W. V. O. (1953/2001) 'Two Dogmas of Empiricism', in *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 20–46.

---

[9] Many thanks to Anthony Fisher for detailed and helpful comments on an earlier version of this chapter. Thanks also to the audience at the Materialist Theory of Mind conference, for enlightening and enriching discussion.

Ryle, Gilbert (1949) *The Concept of Mind*, London: Hutchinson.

Ryle, Gilbert (1970) 'Autobiographical', in *Ryle*, edited by Oscar P. Wood and George Pitcher, New York: Doubleday and Co.

Schlick, Moritz (1925/1974) *General Theory of Knowledge*, translated by Albert E. Blumenber, New York: Springer Verlag.

Schlick, Moritz (1935) 'On the Relation Between Psychological and Physical Concepts', in *Readings in Philosophical Analysis*, edited by H. Feigl and W. Sellars, New York: Appleton-Century Crofts, 393–407.

Sellars, Wilfrid (1956/1997) *Empiricism and the Philosophy of Mind*, with introduction by Richard Rorty, and study guide by Robert Brandom, Cambridge, MA: Harvard University Press.

Sider, Theodore (2011) *Writing the Book of the World*, Oxford: Oxford University Press.

Smart, J. J. C. (1959) 'Sensations and Brain Processes', *The Philosophical Review*, 68(2): 141–56.

Smart, J. J. C. (1967) 'Comments on the Papers', in *The Identity Theory of Mind*, edited by C. F. Presley, St Lucia: University of Queensland Press, 84–93.

Smart, J. J. C. (2000/2017) 'The Mind/Brain Identity Theory', in *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/spr2017/entries/mind-identity/>.

Smith, David Woodruff (1983) 'Is This a Dagger I See Before Me?', *Synthese*, 54(1): 95–114.

Strawson, P. F. (1963) 'Carnap's Views on Constructed Systems Versus Natural Languages in Analytic Philosophy', in *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schilpp, Cambridge: Cambridge University Press, 503–18.

Thomasson, Amie L. (2005) 'First-Person Knowledge in Phenomenology', in *Phenomenology and Philosophy of Mind*, edited by David W. Smith and Amie L. Thomasson, Oxford: Oxford University Press, 115–39.

Thomasson, Amie L. (2017) 'What Can We Do, When We Do Metaphysics?', in *The Cambridge Companion to Philosophical Methodology*, edited by Giuseppina d'Oro and Søren Overgaard, Cambridge: Cambridge University Press, 101–21.

Thomasson, Amie L. (2018) 'Metaphysics and Conceptual Negotiation', *Philosophical Issues*, 27(1): 364–82.

Thomasson, Amie L. (2020) 'A Pragmatic Method for Normative Conceptual Work, in *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen, and David Plunket, Oxford: Oxford University Press: 435–458.

14

Armstrong's Revenge

John Heil

> We must see our cognitive relation to our own mind as like our cognitive relation to anything else in nature. We know in part, guess in part, in part we are mistaken and in a large part we are simply ignorant. Being in a mental state entails nothing about our awareness of that state.
>
> (D. M. Armstrong, *A Materialist Theory of the Mind*, 115)

1. Conservation and Entropy

Philosophers old enough to have read *A Materialist Theory of the Mind* when it was published fifty years ago could be forgiven for harboring the suspicion that philosophy has, in the intervening years, not simply changed, but been in steady decline. The sentiment is likely due in part to a natural compulsion to imagine that things were much better in the old days, but there is more to it than that.

As I recall, 1968, the year in which *A Materialist Theory of the Mind* (hereafter *MTM*) appeared, was also the last year of its being a sellers' market in philosophy—at least in North America—the last year in which there were more jobs in philosophy than applicants. As academic positions in philosophy became increasingly competitive, the pressure to publish encouraged graduate programs to focus on fabricating polished professionals capable of publishing papers on the right topics in the right places, papers calculated not to offend referees looking for reasons to take offense. The upshot has been a proliferation of the philosophical equivalent of elevator music: work that is both inoffensive and deeply uninspiring.

Certainly, it is true that there has always been uninspired philosophy—or if you think, as I do, that philosophy is like art, so uninspired philosophy really is not philosophy—philosophers have always been in short supply.

Two important cosmological principles spring to mind. The first of these I call the Principle of the Conservation of Philosophy. According to this principle, the total amount of philosophy—and remember, I am accepting that uninspired philosophy is not philosophy—remains constant. One result is that when the number of credentialed philosophers increases, philosophy tends to be spread ever more thinly.

John Heil, Armstrong's Revenge In: *Armstrong's Materialist Theory of Mind.* Edited by: Peter R. Anstey and David Braddon-Mitchell, Oxford University Press. © John Heil 2022. DOI: 10.1093/oso/9780192843722.003.0015

A second, and related, principle concerns Philosophical Entropy. Physics tells us that mass-energy is conserved. But physics tells us, as well, that in a closed system entropy increases. So it is in philosophy. Just as the overall amount of philosophy is conserved, closed systems—more or less self-contained domains of research—become progressively impoverished philosophically. We start off with good, strong ideas, then worry them to death. Epicycles are added to epicycles until interesting theses are eventually drained of philosophical life.

In contrast to all this, *MTM* is, as was its author, a force of nature. It is full of important ideas and arguments on a wide range of topics, not simply in the philosophy of mind, but in metaphysics, epistemology, and the philosophy of language. Remarkably, in 376 pages (including a ten-page preface to the second edition), the book includes not a single footnote. This is disciplined, confident, reader-friendly writing at a very high level, and of a kind that, like the quoll in Australia, faculty positions in philosophy, and for that matter, philosophers, is in increasingly short supply.

Enough curmudgeonly flailing about: time to turn to reflections on the book itself.

2. The Concept of Mind

Much of *MTM*—more than half, in fact—is devoted to the purely conceptual project of showing that states of mind are individuated causally. Beliefs and desires, for instance, are distinguished by the fact that they issue from different kinds of cause and produce different kinds of internal and behavioral effect. States of mind are characterized by their distinctive 'causal roles'. A separate, empirically inspired line of argument is meant to establish that the most promising candidates as occupants of these causal roles are states of the nervous system.

The book takes seriously earlier arguments for 'central state materialism' advanced by Ullin Place and J. J. C. Smart in Adelaide. But it also reflects C. B. Martin's—and Max Deutscher's—insistence that mental concepts are 'causally loaded' (Martin 1959, 108–9; Martin and Deutscher 1966). Perceiving a rabbit requires being in causal contact with the rabbit; remembering your tenth birthday party requires a causal connection between the birthday party and your subsequent thoughts about it.

Indeed, despite its title, what was at the time significant about *MTM* was not its advocacy of materialism but its rejection of Wittgensteinian trends previously dominant in the English-speaking philosophical universe. These had been forcefully articulated by Gilbert Ryle and authors of the 'little red books' published in the Studies in Philosophical Psychology series by Routledge & Kegan Paul. (*MTM* was part of another Routledge & Kegan Paul Series, the International Library of Philosophy and Scientific Method.)

The title of my chapter reprises a phrase Armstrong himself uses in the course of examining Elizabeth Anscombe's—'Miss Anscombe's'—account of motives such as revenge (Anscombe 1957). A man's pumping water could be an act of revenge if the man's motives were of the right sort. You might recall Anscombe's example. A man wants to take revenge on the inhabitants of a certain house. The man is pumping because he knows the water he is pumping is poisonous, and he knows that his pumping will supply water to the house. Armstrong calls the man's pumping 'Armstrong's revenge' (*MTM*, 172).

According to Anscombe, ascriptions of intentions and motives to agents are not meant to identify causes of actions; they are, rather, perspicuous redescriptions of actions in light of agents' states of mind. The 'because' in 'the man is pumping because he wants to poison the inhabitants' does not indicate a cause of the man's action. Intentions and motives explain actions by allowing us to recast those actions in an illuminating way, a way that enables us to make sense of them: the man is pumping, the man is supplying water to the house, the man is poisoning the inhabitants.

So Armstrong's targets are not primarily defenders of dualism, but behaviorists and anti-causalists who took their cues from Wittgenstein. Your acting for reasons, for instance, was not, as the anti-causalists held, simply acting 'in light of' your beliefs and desires, but being caused to act by your beliefs and desires. Similarly, your perceiving a tomato rolling down the supermarket aisle was a matter of your being caused to believe that there is a tomato rolling down the supermarket aisle by a tomato's rolling down the supermarket aisle. Reliably caused beliefs—beliefs caused by information-conveying perceptual mechanisms—were thereby justified, their justification stemming from their causal histories. Beliefs based on these beliefs were mediately justified, the basing relation being understood causally.

Although the epistemology is important—and, I believe, broadly correct—my chief concern here is Armstrong's functionalist conception of the mind. The label is appropriate, although Armstrong himself does not use it, and Armstrong's version of functionalism differs importantly from what came to be regarded as 'mainstream' functionalism.

3. Two Strains of Functionalism

In an influential paper published in 1980, Ned Block distinguished two functionalist camps. First, there were those who, like Armstrong and David Lewis (1966), regarded mental predicates or concepts as what Block calls 'functional specifiers'. Mental states are states that play causal roles associated with particular mental concepts.

Suppose that being in pain is a matter of being in a bodily state brought about by tissue damage that issues in characteristic ameliorative 'pain behavior', and suppose that a particular occurrence in your nervous system satisfies this description on a given occasion. Then, for Armstrong, as for Lewis, that state *is* your being in pain.

In contrast, a second version of functionalism—what Block somewhat confusingly calls the 'functional state identity' version—identifies your pain, not with the state that 'plays the pain role', but with your having the property of being in such a state. Pain is a 'higher-order' property, where a higher-order property is understood as a property you have by virtue of having a property of a particular kind. Sometimes this is put as the view that pains are not the occupants of the pain role, but the 'role itself'. States playing the pain role are not themselves pains, but 'realizers' of pains.

This strange doctrine is a principal source of what came to be known as 'nonreductive physicalism', a conception of mental states as 'higher-level' states, dependent on, yet distinct from their physical realizers. You are in a state that plays the pain role and, by virtue of being in this state, by virtue of having a particular physical property, you have the property of being in pain. This is what became mainstream functionalism in North America, the wellspring of the problem of 'causal relevance'.

4. Causal Relevance

Your grasping the handle of a hot frying pan results in a painful burning sensation, and you quickly withdraw your hand. It might seem that the painful feeling is causally responsible for your behavior. Mainstream—functional state identity—versions of functionalism call this picture into question. It is not the pain, not the mental property, but the realizer of the mental property that is doing the causal work. The mental property, at best, goes along for the ride. In general, mental properties lack causal relevance, their causal efficacy is pre-empted or 'screened off' by their physical realizers.

Just when we thought we had once and for all put dualism behind us, we find ourselves saddled with a new variety of dualism: property dualism. Mental properties are a species of higher-order—or, better, higher-level—property, dependent on, but distinct from, lower-level physical realizing properties. And given that the causal story in which these lower-level realizers figure is complete, there is nothing for the mental properties to contribute. Mental properties must be either epiphenomenal or, implausibly, 'overdetermine' physical effects.

This problem, if it is a problem, does not arise for Armstrong: your mental states and processes *are* those states and processes causally responsible for your behavior. But, in that case, what happens to the evident 'multiple realizability' of the mental?

One way to understand Armstrong's position requires distinguishing mental predicates and mental properties or states. (The assumption here is that a state is something's having a property.) A mental predicate expresses a mental concept satisfied by properties or states individuated causally. What mainstream functionalists see as the realizers of mental properties would be, for Armstrong, truthmakers for applications of mental predicates. Where mainstream functionalists wax lyrical about the multiple realizability of mental properties, Armstrong sees mental predicates as being multiply *satisfiable*, that is, as being true of different, but broadly similar, kinds of property, only provided the properties have the right causal profile.

I am putting this in my own words, not Armstrong's, but I believe this captures the distinction between mainstream functionalism and the version favored by Armstrong and by Lewis. The tragedy is that philosophers of mind by and large ignored Armstrong–Lewis functionalism in favor of the 'functional identity' view, following Block, Sydney Shoemaker, and other mainstream functionalists down the rabbit-hole of nonreductive physicalism (see, for instance, Shoemaker 1984).

5. One Over Many

You might think that mainstream 'functional identity' functionalism enjoys an important advantage over the 'functional specifier' version favored by Armstrong. For Armstrong, your being in pain is your having a property—your being in a physical state—with the right sort of causal profile. These physical states could be dramatically different across species. For the mainstream functionalist, in contrast, your being in pain is not your being in one of these states but your having the property of being in one of these states, your having the property of having a property with the right causal profile. And this is something shared by every creature in pain, regardless of that creature's physical makeup.

So mainstream functionalism provides unity where Armstrong has diversity. This would seem to be an important explanatory benefit, one that allows for lawful generalizations about pains that would elude Armstrong.

What is it to have the property of having a property? If you have a certain property, it is true that you have it. But is it true that you thereby have the property of having a property? What would this higher-order property contribute? Is it required for unity? I have suggested that one way to express the difference between mainstream functionalism and Armstrong's functionalism would be to say that mainstream functionalists regard mental properties as multiply realizable whereas Armstrong regards mental predicates—'is in pain', for instance—to be multiply satisfiable. The pain predicate can apply truly to very different kinds of creature, provided only that these creatures possess similar properties, properties with similar causal profiles.

The unity afforded by the higher-order property is a kind of cheesy philosophers' unity. You collect similar properties and invent the property of having any of these (or any property sufficiently similar to these). A position of this kind enjoys no advantage over Armstrong's. The pain predicate encompasses causal similarities, the same similarities required to generate the spurious unifying higher-order property.

This is not simply a stand-off, however. By making mental properties higher-order properties, mainstream functionalists visit upon themselves the problem of causal relevance. There is nothing for the higher-order properties to do. Any causal contribution they might make is preempted by their realizers. So, ironically, a causal theory of the mental takes the mental out of the causal mix.

6. Supervenience

In addition to multiple realizability, another factor contributing to the rise of nonreductive physicalism was the publication, in 1970, of Donald Davidson's 'Mental Events' (Davidson 1970). Davidson argued that the possibility of mental causation rested on the 'supervenience' of the mental on the physical: mental events—which, for Davidson, were events describable in a mental vocabulary—'supervened' on physical events—events describable in a non-mental, physical vocabulary. Davidson's supervenience thesis was the thesis that every mental event was—that is, was identical with—a physical event. This, really, is just a variation on the Armstrong view. Mental terms pick out states of the world that happen to be physical: mental predicates designate states satisfying physical predicates.

In the 1970s and 1980s, philosophers besotted with the idea that mental properties were higher-level, multiply realized properties pounced on the notion of supervenience, construing supervenience not as a contingent relation among ways of describing the universe, but as a robust metaphysical relation among 'families of properties' (see Kim 1984; 1993). Davidson was a staunch anti-reductionist in the straightforward sense that he did not think you could analyze or reformulate mental predicates or descriptions in a non-mental vocabulary. When you combine anti-reductionism and a commitment to the supervenience of the mental on the physical with the idea that supervenience is a relation among families of properties, you get nonreductive physicalism.

Nonreductive physicalism could be seen as the embarrassing progeny of two unfortunate but avoidable philosophical accidents: the ascendency of mainstream functionalism over Armstrong–Lewis functionalism, and the embrace of supervenience as a thesis about necessitation relations among distinct kinds of property. Both, however, have their roots in philosophers' complicity in arguments that begin with linguistic premises concerning predicates and taxonomies, and end with metaphysical conclusions concerning relations among properties.

Davidson's position was, in fact, much closer to Armstrong's than it was to Kim's. For Davidson, every event picked out by a mental predicate or description, also satisfies—that is, makes true—some physical description, where you could think of a physical description as a description couched in the vocabulary of fundamental physics. Armstrong differs from Davidson in his willingness to countenance the possibility that a state picked out by a mental predicate could fail to satisfy a physical predicate.

You might regard all this as little more than a historical curiosity—if indeed you do not reject my account outright. You might, for instance, find functionalism hopeless owing to functionalism's presumed inability to accommodate qualities of conscious experiences, the so-called *qualia*. I would like to persuade you that the kind of functionalism I have ascribed to Armstrong is not, despite appearances, susceptible to this worry, and that it is in fact well positioned to cope with conscious experiences.

Before explaining how this could be so, however, I must address what I see as a difficulty with Armstrong's treatment of dispositions. Although I believe Armstrong mishandles dispositions, the difficulty in question is easily correctable. Correcting the matter will open the way to answering the worry about the incompatibility of functionalism and conscious qualities.

7. Causality and Dispositionality

Powers (or dispositions—I will use the terms interchangeably) have been central to the ways scientists, philosophers, and everyday observers have thought about the universe from the beginning. Aristotle and his medieval successors invoked powers to explain the behavior of inanimate objects and organisms. Newton and Newton's philosophical ally, Locke, continued the tradition, but it had come to be challenged by scholastics concerned to emphasize God's omnipotence, and then, more aggressively in the seventeenth century, by the Cartesians. For Descartes, material bodies do what they do, not because they are as they are, but because they are governed by contingent laws of nature. These laws were taken by Descartes to be expressions of principles freely chosen by God. The tomato rolls, not because it is spherical and thereby has the power to roll, but because God ordains and thereby ensures that spherical things roll.

Descartes thus strips powers from the objects and relocates them in God. On such a conception, God, and perhaps finite minds, are the exclusive source of efficient causation in the material world. Material bodies merely transmit motions, the origin of which lies outside the material realm. Hume adopts the Cartesian picture minus God. Removing God from the scene removes any sense in which material bodies could be said to be *governed* by laws. At most, laws express exceptionless generalizations concerning the actual behavior of objects.

This Humean picture dominated analytical philosophy in the twentieth century and, to an extent, still does. It was occasionally challenged—by C. B. Martin in Sydney, for instance, and by Harré and Madden (1975)—but, by and large, it became the default.

One difficulty with the Humean program stems from the fact that we are interested not only in what things do, but in what they *would* do. Hume himself understood beliefs about such matters—typically expressed in counterfactual and subjunctive conditionals—as observers' projections of their own conditioned expectations. Such a move imports an untoward subjectivity into the scientific enterprise, but it also forces us to look more closely at observers. Evidently there are truths about observers based not only on what they do, but on what they would do, what they are *disposed* to do as a consequence of conditioning.

Ryle says this clearly, but then rests content with a broadly Humean framework according to which counterfactual and subjunctive conditionals were understood, not as designating causally productive states of agents, but as 'inference tickets'. Others were less inhibited, accepting that minds are *affected* by what is perceived, and *issue* in behavior. This seems to make states of mind causal—or better, dispositional, where dispositions are regarded with full ontological seriousness. Thus, the emergence of functionalism and attendant causal accounts of action, perception, and knowledge.

A Materialist Theory of the Mind occupies a central place in the rejection of behaviorism and the development of causal–dispositional accounts of states of mind. However, while philosophers seemed ready to abandon Humeanism when it came to the mind, Humeanism continued to dominate elsewhere, especially in metaphysics, including the metaphysics of causation. One result was a kind of repressed tension in the way philosophers—Armstrong included—came to think about causality and dispositionality.

8. Dispositions

In *The Concept of Mind*, Ryle appeals to dispositions in the course of explicating various mental concepts. To believe that the ice is thin is, according to Ryle, to be disposed to behave in various ways—to avoid skating on the ice, for instance, to warn others not to skate, to fail to express surprise when someone does skate and falls through the ice. Ryle, however, is not a realist about dispositions: 'To possess a dispositional property is not to be in a particular state, or to undergo a particular change; it is to be bound or liable to be in a particular state, or to undergo a particular change, when a particular condition is realised' (Ryle 1949, 43).

Armstrong, in contrast, takes dispositions ontologically seriously, embracing the thesis that dispositions have nondispositional categorical 'bases'. He contrasts this with 'Phenomenalist or Operationalist' conceptions such as Ryle's (*MTM*, 87):

> To this we may oppose . . . a Realist account of dispositions. According to the Realist view, to speak of an object's having a dispositional property entails that the object is in some non-dispositional state or that it has some property (there exists a 'categorical basis') which is responsible for the object manifesting certain behaviour in certain circumstances, manifestations whose nature makes the dispositional property the particular dispositional property it is. It is true that we may not know anything of the nature of the non-dispositional state. But, the Realist view asserts, in asserting that a certain piece of glass is brittle, for instance, we are *ipso facto* asserting that it is in a certain non-dispositional state which disposes it to shatter and fly apart in a wide variety of circumstances. (*MTM*, 86)

Earlier I invoked Block's distinction between two strains of functionalism, noting that Armstrong's brand of functionalism identifies mental states with whatever states of an agent play the causal roles characteristic of states of that kind. This is what Block calls the 'functional specifier' version of functionalism, distinguishing it from a more influential variant, according to which mental states and properties are identified with higher-order or higher-level states and properties. On this conception, functional states generally and mental states in particular are not the occupants of functional roles, not the states that as a matter of fact play the roles, but (as it is sometimes put) the roles themselves. Thus, a mental state is not a state that plays a particular causal role but a state you are in by virtue of being in a state that plays that role. If this strikes you as confusing, you are not alone. And yet it is something philosophers have swallowed with barely a peep since the advent of functionalism decades ago.

On Armstrong's official view, a mental state could be a brain state with the right causal–dispositional profile. But this conception is hard to square with what Armstrong says about dispositions. Armstrong's dispositions closely resemble the mainstream functionalist's mental states. They are states something is in by virtue of being in some *non*dispositional, categorical state.

Now, if you take this view of dispositions—essentially a 'functional state identity' account of dispositionality of the sort artfully defended in Prior, Pargetter, and Jackson (1982)—you inherit all the 'causal relevance' problems associated with mainstream functionalist accounts of mental states: what does the causal work is not the disposition but its categorical 'ground' or 'basis', its nondispositional realizer. Frank Jackson, for instance, regards the thought that dispositional properties might themselves be causally relevant as 'a curious and ontologically extravagant kind of overdetermination' (1998: 202).

Here is the state of play. Armstrong accepts a 'functional specifier' account of mental states, according to which states of mind could be identified with brain states with the right causal–dispositional profiles. At the same time, he accepts a 'functional identity' account of dispositions, according to which dispositions are not identifiable with states of the brain, which turn out to be nondispositional, categorical states that realize dispositions. The states themselves are not dispositional but merely the 'bases' of dispositions.

The inherent instability of this position could be resolved in either of two ways. First, Armstrong could turn his back on 'functional specifier' functionalism and throw his lot in with mainstream 'functional identity' functionalism. On such a conception, mental states are realized by, but not identified with, causal-dispositional states that are themselves are realized by, but not identified with, categorical states of agents. Put in terms of properties, mental properties supervene on dispositional properties that in turn supervene on nondispositional properties.

This option would locate mental states at two removes from states taken to be doing the 'real causal work'. The result would be a conception of mental states and properties that makes them not simply epiphenomena, but epiphenomenal epiphenomena. It is hard to see how this result could be avoided if you accept the idea that dispositions must have nondispositional bases, categorical realizers. The very same arguments get you to mainstream functionalism about mental states.

A second option is, in my opinion, much more attractive and closer to the spirit of the kind of plain-spoken Australian realism exemplified by Armstrong. This option jettisons the requirement that dispositions have nondispositional bases, in effect, adopting a 'functional specifier' conception of dispositions. Now states of mind are identified with dispositional states and these are identified with categorical brain states.

A conception of this kind amounts to a repudiation of the thesis that categorical states are nondispositional. If this strikes you as paradoxical, perhaps that is because you accept the idea that 'categorical' *means* 'nondispositional' (and 'dispositional' means 'noncategorical'). But why think that? Ryle, an anti-realist about dispositionality, thought it important to distinguish something's here and now 'categorical' features from its 'noncategorical', merely 'if–then' features, but there is no reason to follow Ryle in this. If Jackson is right and the categorical bases are what do the causal work, why not simply regard the *bases* as the dispositions? Indeed, this is precisely what Armstrong appears to suggest in the last sentence of the passage quoted earlier: 'in asserting that a certain piece of glass is brittle, for instance, we are *ipso facto* asserting that it is in a certain non-dispositional state that disposes it to shatter and fly apart in a wide variety of circumstances'.

Realism about dispositions makes them fully here and now, fully categorical states of whatever possesses them. These are what answer to dispositional predicates. Many different kinds of categorical state or property can answer to a single dispositional predicate, hence the way is open for a 'functional specifier' account of dispositional states. All in all, such a realist conception of dispositions seems not only vastly preferable to the alternative, but, more to the point, a much better fit for Armstrong's brand of functionalism.

Peter Anstey suggested to me that this is at odds with Armstrong's antipathy toward dispositional properties (see, for instance, Armstrong 1997: 75–80). Armstrong's dispositions must have 'categorical bases'. The kind of realism I am attributing to Armstrong, however, accepts that properties are uniformly categorical. A property would be dispositional, not by virtue of being non-categorical, but by virtue of standing in relations to other properties, relations that constitute causal laws. This is what the 'categorical basis' of a disposition amounts to. Armstrong's brand of functionalism requires only first-order properties standing in lawful relations. Epiphenomenalism is not on the cards.

9. Qualities

Having cleared the way for an account of dispositional states that comports with Armstrong's functionalist conception of the mind, we are left with a problem that has plagued functionalism from the start: the problem of the qualitative nature of conscious experiences (see, e.g., Campbell 1970). Functionalists individuate states of mind by reference to their causal–dispositional profiles. But this appears to omit what is most immediately striking about conscious states: their distinctive qualitative character.

Once philosophers became comfortable with the idea that mental concepts include essential causal elements, they fell into the trap of concluding that *all there is* to a mental state or property is its causal–dispositional nature. This is analogous to the increasingly popular idea that all there is to the properties of material bodies is their causal–dispositional profile: material properties are 'pure—that is, *purely*—powers'. If you couple this with the confused idea that because powers are individuated by reference to ways they manifest themselves in concert with other powers, powers must be 'relational properties', you risk succumbing to a ratbaggish 'structural realism', according to which relations are fundamental, their relata derivative.

The dialectic is illuminating. You start with all-out Humeanism, according to which all that exists is a four-dimensional spread of qualities: dispositionality is deeply mysterious. You then graduate to a conception of properties as powers and regard qualities with suspicion. One option would be to allow that properties of material bodies are powers, and relegate qualities to conscious experiences. Although this option might appeal to physicists with no interest in qualities, it has the effect of bifurcating the universe into distinct material and immaterial domains. The initial appeal of functionalism lay in its rejection of this kind of dualism. Minds could be seen as dynamic material systems. The trouble is that functionalism appears to ignore what is most salient about conscious experiences: their intrinsic qualitative character.

My suggestion is that, just as it is a mistake to see the categorical and dispositional as mutually exclusive categories of property, so it is a mistake—the same mistake, really—to see the dispositional and the qualitative as mutually exclusive categories of property. Dispositional properties *are* categorical, *are* here and now, even if their manifestations are not. Qualities *are* dispositional, objects are powerful owing to their qualities.

A tomato's sphericality and color are both qualities of the tomato. But it is by virtue of being spherical that the tomato would roll, would make a circular concave impression in a pillow, would look and feel spherical. Similarly we can see the tomato's color because its color—its redness—affects light radiation so as to arouse experiences of redness in suitably equipped observers. Changing the qualities changes the powers. The distinction between qualities and powers is one of conception only—more specifically it is what the scholastics and early moderns called a *distinction of reasoned reason*. You can consider something's powers apart from considering its qualities, and you can consider its qualities apart from its powers, but you cannot prise these apart in reality.

C. B. Martin inflicted this position on me years ago, and it has by now become second nature. The position comports with both common sense and scientific practice. Science abstracts from the qualitative nature of objects because the aim is to explain and predict how objects behave or would behave in particular circumstances. It would be a mistake to conclude from this that the objects lack qualities, however. Cooks and audiophiles know that the qualities of ingredients determine how they would taste or sound.

10. Conscious Qualities

Pretend for a moment that there is something in this and return to the kind of functionalist picture painted in *MTM*. There, states of mind and mental properties are individuated by their dispositional nature. But why not think that a particular property is apt for a particular role—the pain role, for instance—because it is as it is qualitatively? Remember, I am denying that qualitative and dispositional natures could vary independently. Far from excluding qualities, Armstrong's functionalism would allow them to play a pivotal role in the operation of minds.

You might find this suggestion utterly implausible. If Armstrong is right, states of mind, including conscious states, are to be identified with states exhibiting the right causal–dispositional profiles. In the case of terrestrial creatures, we have good reason to think that the pertinent states are states of the central nervous system: brain states. Is it plausible, however, to suppose that the qualities associated with conscious experiences could be qualities of brains or brain processes? Imagine having a visual experience of a ripe tomato or feeling a burning pain in your biceps. A scientist inspecting your brain might discover neuronal goings on *correlated* with these experiences, and these neuronal goings on might have assorted qualities, but these qualities are nothing at all like the qualities of your experiences.

This is less than perfectly obvious, however. First, consider your visual experience of a ripe tomato, and distinguish qualities of that experience from qualities of what you are experiencing. The tomato is red and spherical, your experience of the tomato is an experience of something red and spherical, but your *experience* is not red and spherical. It is scarcely surprising, then, that a neuroscientist examining your brain when you are undergoing this experience will not encounter anything red and spherical. (Identifying qualities of objects experienced with qualities of experiences is what Ullin Place called the 'phenomenological fallacy'; see Place 1956; Smart 1959.)

Second, it is important to understand that what is to be compared are your visual experiences of a ripe tomato and a neuroscientist's experiences of your experiences, which, it is being supposed, could be occurrences in your brain. There is no reason to think that these experiences should be alike qualitatively. Just so, there is no reason to think that the neuroscientist could work out what it is like for you to be in your experiential state from observations—experiences—of that state.

I am assuming here that your 'knowing what it is like' to undergo a particular kind of experience is intimately related to your undergoing (or having undergone) experiences of that kind. And again, to the extent that an experience of a ripe tomato differs from the experience of an experience of a ripe tomato, a neuroscientist's undergoing an experience of your visual experience of a ripe tomato need not resemble what it is like for you to undergo that experience.

What of pains and other 'bodily sensations'? It is all well and good—or at least not entirely outrageous—to locate the redness and sphericality you experience when you have a visual experience of a ripe tomato in the tomato. But where do you locate the qualities of pains? Armstrong thinks of bodily sensations as agents' experiences of their own bodily states. Suppose that when you are in pain on a particular occasion, you are experiencing a muscle contusion, and suppose you experience this as a burning sensation. It is no good locating the painful qualities in the muscle contusion. Or at least doing so would require a heroic philosophical move not called for in locating experienced redness and sphericality in a tomato.

Unlike the tomato case, if your bodily sensations are experiences of your bodily states, the pertinent qualities are qualities of the experiences, not qualities of the bodily states your experiences are experiences of. You can have an experience of a tomato in the absence of a tomato—as when you hallucinate a tomato—and you can have an experience of pain in your biceps without a contusion in your biceps—as when you experience a referred pain or a phantom pain—but you cannot have an experience of pain without the pain. It is the experience, not the bodily state that produces the experience, that is identified with the pain, it is the experience that is painful.

Of course, if Armstrong is right, this experience is itself—or at any rate could be—a state of your nervous system. If this were so, and if your experience had certain qualities, then this state must have these qualities. Is this a credible position?

Perhaps it is. To see why, consider again the difference between something's being in a given state and your observing its being in that state. Your observing the state is a matter of *your* being in a particular experiential state. And, as before, there is no reason to think that your observational state need resemble the state being observed—or, for that matter, to reveal what, if anything, it is like to be in that state. If you allow this, the way is open to locate qualities of bodily sensations in the nervous systems of experiencing creatures. And why not?

Of course, if you start with the assumption that the qualities in question could not be qualities of something like a nervous system, then this will seem unlikely. But where does this assumption come from? What is to be said for it? Once you distinguish between (1) your being in an experiential state and (2) a neuroscientist's being in a state of experiencing your experiential state, the vague thought that a neuroscientist's observing your being in pain must put the neuroscientist in a state in some fashion resembling yours loses traction.

11. Armstrong Redux

More could be said here, of course, but I hope I have at least managed to cast doubt on the idea that experiential qualities could not possibly be qualities of brains. I believe that this is at least consistent with the conception of states of mind, including bodily sensations Armstrong advances in *A Materialist Theory of the Mind*. The beauty of that conception is that it reconciles a broadly functionalist picture of the mind with the evident importance of qualities of conscious experiences. By my lights, Armstrong was miles ahead of the functionalists and the anti-functionalist '*qualia* freaks' that succeeded him. *A Materialist Theory of the Mind* is a philosophical treasure that, even more than fifty years on, repays close reading.

References

Anscombe, G. E. M. (1957) *Intention*, Oxford: Basil Blackwell.

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge & Kegan Paul.

Armstrong, D. M. (1997) *A World of States of Affairs*, Cambridge: Cambridge University Press.

Armstrong, D. M. (1993) 'Preface', in *A Materialist Theory of the Mind*, 2nd edition, London: Routledge, xiii–xxiii.

Block, Ned (1980) 'What Is Functionalism?', in *Readings in Philosophy of Psychology*, vol. 1, edited by Ned Block, Cambridge, MA: Harvard University Press, 171–84.

Campbell, K. (1970) *Body and Mind*, Garden City, NY: Doubleday and Co.

Chalmers, David J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, New York: Oxford University Press.

Davidson, Donald (1970) 'Mental Events', in *Experience and Theory*, edited by Lawrence Foster and J. W. Swanson, Amherst, MA: University of Massachusetts Press, 79–101.

Harré, Rom and E. H. Madden (1975) *Causal Powers: A Theory of Natural Necessity*, Oxford: Basil Blackwell.

Jackson, Frank (1998) *From Metaphysics to Ethics: A Defense of Conceptual Analysis*, Oxford: Clarendon Press.

Kim, Jaegwon (1984) 'Concepts of Supervenience', *Philosophy and Phenomenological Research*, 45: 153–76.

Kim, Jaegwon (1993) *Supervenience and Mind: Selected Philosophical Essays*, Cambridge: Cambridge University Press.

Lewis, David (1966) 'An Argument for the Identity Theory', *Journal of Philosophy*, 63: 17–25.

Martin, C. B. (1959) *Religious Belief*, Ithaca, NY: Cornell University Press.

Martin, C. B. (1994) 'Dispositions and Conditionals', *Philosophical Quarterly*, 44: 1–8.

Martin, C. B and Max Deutscher (1966) 'Remembering', *The Philosophical Review*, 75: 161–96.

Place, U. T. (1956) 'Is Consciousness a Brain Process?', *British Journal of Psychology*, 47: 44–50.

Prior, Elizabeth W., Robert Pargetter, and Frank Jackson, (1982) 'Three Theses about Dispositions', *American Philosophical Quarterly*, 19: 251–7.

Ryle, Gilbert (1949) *The Concept of Mind*, London: Hutchinson.

Shoemaker, Sydney (1984) *Identity, Cause, and Mind*, Cambridge: Cambridge University Press.

Smart, J. J. C. (1959) 'Sensations and Brain Processes', *The Philosophical Review*, 68: 141–56.

Index

Adaptationism about the mind 179–80
Alexander, S. 33, 34 n8
Alter, T. 214
Anderson, J. 2, 33–5, 55, 61
Andrews, K. 235
Anscombe, G. E. M. 206, 246
Anstey, P. R. 1–21 *passim*, 254
Appearances, *see* seemings
Aristotle 250
Armstrong, D. M.
  Anderson's influence on 2, 33–5
  on authority of science 9, 10–11, 15, 36–7, 227, 231
  BPhil thesis 34, 63, 65
  and conceptual analysis 1, 15–16, 27, 29, 35–6, 56, 77, 215, 229, 230, 233, 240, 241
  his copy of Ryle's *The Concept of Mind* 17
  and DNA analogy 13, 67
  and functionalism 7, 8, 12, 33, 76, 211, 214, 215, 225, 246–55
  headless woman illusion 11
  and history of Causal theory 2, 12–13, 16, 26–7
  and the Identity theory 1, 12, 13–14, 24 n1, 29, 33, 58, 67, 69, 76, 232
  immediate reception of *MTM* 195
  C. B. Martin's influence on 12 n4, 17–18, 37, 65–6, 71, 245
  methodological principles of 2, 9, 15, 18, 39–40
  and Oxford 15, 16, 34, 56–60, 229
  H. H. Price's influence on 34–5, 37
  and 'the split' in 1973 76–7
  study group on space and time 55–6
  as a teacher 76
  and type-type identity 13, 14
  undergraduate and honours papers 2, 10, 53–5
  see also, Ryle
Armstrong, D. M. (Works)
  A Combinatorial Theory of Possibility 76
  Belief, Truth and Knowledge 83, 207
  Berkeley's Theory of Vision 10, 34, 61
  Bodily Sensations 10, 17 n8, 35, 58, 62, 63, 96 n4, 105, 216–17
  Consciousness and Causality 145 n6, 181
  'Epistemological foundations' 11, 12, 15
  'Is introspective knowledge incorrigible?' 147, 149, 150, 157–9
  Perception and the Physical World 10, 35, 62, 65, 66, 71
  'Self-Profile' 11 n1, 13, 15, 62, 69
  The Mind-Body Problem 12, 128
  'The causal theory of the mind' 12, 15, 197
  'The nature of mind' 10, 11, 13, 16, 17 n8, 36, 64, 67, 68
  Universals and Scientific Realism 35 n9
Austin, J. L. 40
Ayer, A. J. 32, 40, 60, 66

Bach, K. 203 n13
Baier, K. 147 n16
Barron, A. B. 177 n1
Behaviourism 1, 2, 11, 19, 26, 27, 30–3, 38, 41, 44, 45, 47, 77, 78–80, 251
  Carnapian 30, 44–5
  Positivist 38, 44–5
  Rylean 15–17, 31, 38, 39, 40, 44, 45, 46, 78, 79, 195–6, 198 n6
Belief 2, 3, 5, 21, 66–7, 79–80, 82, 97, 100, 102, 105, 169–71, 173, 176, 178, 185, 187, 189, 196, 201, 202, 205, 207, 228, 234, 235–6, 237, 240, 245, 246
  as causes 79, 176, 205, 245–6
  as dispositions 66, 67, 93
  illusory 37
  introspective 143, 145, 157, 165, 168, 169, 170, 172, 191–2
  as maps 3, 82–3
  moral 200–1
  perception as acquisition of 3, 35 n10, 37, 92–7, 100, 104, 106, 114–16, 128, 176, 184, 238, 241
  perceptual 82, 101, 102, 144, 156, 169
  representational role of 100
  tacit 206
  see also, introspection, perception
Belief-desire psychology 78–84, 196, 199 n8, n9
  belief-desire nexus 2, 78, 79, 83
  belief-desire profiles 3, 80, 81–3
Berkeley, G. 34–6, 60–1, 78, 130, 131
Blindsight 109
Block, N. 7, 181, 182 n7, 246–8, 252
Boeckx, C. 186
Boghossian, P. 120 n11, 214 n3
Bourne, C. 218

Braddon-Mitchell, D. 1–21 *passim*, 82 n11, 190 n13, 211–25 *passim*
Bradley, B. 203, 205
Bradley, F. H. 61
Bradley, M. C. 41, 232
Braithwaite, R. B. 82 n10, 197 n4
Bratman, M. 199 n9
Breitmeyer, B. G. 109
Brentano, F. 143, 144, 197
Byrne, A. 117 n6, 118, 122 n12

Campbell, K. 19, 211, 254
Carnap, R. 19, 30, 32, 44–6, 63–4, 82 n10, 231–2, 234, 235
  see also, behaviourism
Carruthers, P. 146 n10
Categorical properties, *see* properties
Causal theory of the mind 2, 3, 4, 6, 7, 11–14, 16, 21, 77, 79, 94–6, 99, 100, 125, 126, 127, 132, 133, 207 n23, 211, 213, 214, 215, 225, 237, 249
  see also, functionalism
Causation 1, 2, 6, 9, 10, 11–12, 14, 18, 24, 26–7, 37–8, 41, 52–4, 56, 58, 62, 68, 69, 71–2, 173, 211, 214, 216–22, 230, 231, 250–1
  causal conditionalism 211–25 *passim*
  causal relevance 247–9, 252
  causal role 8, 12, 14, 15, 21, 26–7, 41, 43, 68, 69, 72, 77, 79, 82, 127, 133, 137, 164, 170, 171, 212–13, 218, 223, 224, 225, 245, 246, 247, 252
  counterfactual theory of 24, 251
  mental 7, 176, 177, 178, 189, 196, 197, 205, 249
  non-physical 14, 215–19
Chalmers, D. J. 19, 85 n17, 88, 97, 156–7, 208 n25, 211
Chappell, V. 195
Chisholm, R. M. 79, 83 n12, 198 n6, 205, 206
Choi, S. 64
Clark, A. 99 n5, 136, 137
Colour 3–4, 34, 42, 62, 93, 94–6, 98, 99, 100, 101, 102, 112, 116–23, 127–31, 132, 133, 136, 139, 140, 154–6, 159, 162, 237, 255, 256
  colour solid 120
  constancy 118, 120
  inversion 128, 136, 139
  realism 3, 112, 118–22
  reflectance 3, 93, 118–23
  subjectivism 119
  true blue 122
Conation 5–6, 195–206 *passim*
  see also, desire
Conceivability argument against materialism 6, 19, 211–25 *passim*
Conceptual analysis 1, 6, 15–16, 24, 27, 29, 35, 36–8, 56, 77, 215, 217, 227–30, 233, 240, 241, 245
Conceptual ethics 6, 230–2, 236, 240–2
Consciousness 3, 5, 9, 29, 31, 35, 37, 87, 96–100, 103–4, 105–9, 118 n7, 125, 128, 132, 134, 137–41, 144 n2, 161 n33, 176–93 *passim*, 211–19, 223, 250, 254, 255–7
  and biological function 177–83, 184
  higher-order theory 106–9, 182
  perceptual 181, 184–7
  pluralism about 181–2
  utility of 108–9
  see also, introspection
Content externalism 206, 208
Cook Wilson, J. 34
Cox, R. 4, 5, 143–73 *passim*
Crane, T. 71, 137
Crawford, S. 30, 31, 45, 232

Davidson, D. 196 n2, 249–50
Dennett, D. C. 76, 84 n14, 196 n2, 199 n8
Deontic logic 203 n15, 204
  deontic paradoxes 205
Department of Traditional & Modern Philosophy 76, 77, 206 n22
Descartes, R. 63, 85, 128, 250
Desire 5, 6, 21, 38 n12, 78–84, 109 n11, 166 n39, 171, 176, 178–9, 185, 189, 195–205 *passim*, 228, 235, 236, 245, 246
  conditional 203–5
  'Grain' problem for 5, 202–3
  intentionality of 207–8
  pro-attitudes 198–201
  proper 199–201
  see also, belief, conation, hopes, wishes
Devitt, M. 76, 206, n22
Deutscher, M. 37 n11, 245
Dewey, J. 88, 89 n21
Dienes, Z. 98, 104
Dispositions 2, 3, 7, 9, 11, 12 n5, 16–18, 30, 37–9, 52, 56, 58–60, 63–72, 79, 80, 81, 89, 93, 195–6, 229 n3, 250–5
  Chisholm-Geach objection 79
  Rylean 16, 65, 72, 79, 195–6, 229 n3, 251–3
  see also, properties
Distinctness (ontological) 5, 132, 143–73 *passim*
  argument from distinctness 150–2
  of body and mind 127
  constitutive 155–162, 168–73
  distinct existences argument 145
  numerical 152–5, 156, 157, 159, 161–2, 163–5, 170, 171, 172
  see also, Hume (Distinctness principle)
Dorsal and ventral visual stream 139–40
Dretske, F. 76, 146

Dualism 10, 12, 18–19, 21, 27, 31, 42–3, 85–8, 147, 161, 213–15, 222, 224, 225, 232–4, 246, 254–5
  Cartesian 63, 85, 231
  property 32 n2, 247

Eliminitivism about the mind 46, 229 n3, 233–5, 241
Elliott, E. 81, 82 n9
Empiricism 31, 78
Epiphenomenalism 42, 86, 138, 247, 253–4
Experience 4, 18–19, 21, 25, 26, 30, 36, 38, 42, 43, 57, 87–8, 126, 156, 213 n2, 217, 219, 238, 239, 240
  ascriptions 41, 43
  conscious 125, 132, 134, 137–9, 150, 250, 254, 255, 257
  perceptual 3, 113–23, 128–9, 131, 135, 136, 140, 141, 237, 256, 257
Explanatory gap argument 4, 125–41
Expressivism about sensation statements 31, 237–39

Fara, D. G. 202–3
Fara, M. 64
Farkas, K. 1, 4, 125–41 *passim*
Feigl, H. 2, 13, 19 n11, 25 n2, 30–2, 46, 47, 232
Feldman, F. 205
Feyerabend, P. K. 229 n3, 232–4
Fisher, A. R. J. 2, 17 n8, 29–47 *passim*
Flanagan, O. W. 177 n1
Fodor, J. 76
Ford, S. 2, 52–72 *passim*
Folk psychology 29, 43, 45, 47, 116, 164
Forber, P. 179 n3
Forbes, G. A. 218
Forrest, P. 218, 219
Foster, J. 115 n4
Function 5, 135, 179–83, 188, 190, 191, 193, 228, 229 n3, 235–7, 239, 240, 241
  biological 5, 176–180, 183, 184, 187
  functional definitions 29, 43, 45
  functional duplicates 181 n6, 215–16
  representational 100
  see also, consciousness, causation, functionalism
Functionalism 7–8, 21, 76, 138, 246, 251, 254, 255
  analytic 6, 7–8, 33, 43, 211–25 *passim*
  causal 12, 14
  empirical 222
  functional specifier 7, 246–50, 252–5
  functional state identity (mainstream) 7, 247–9, 252, 253
  phenomenal 126, 136–8
  realizer 7, 8
  role 7, 248

Galileo 130
Geach, P. 79, 81 n9, 83 n12, 198 n6
Gertler, B. 146 n10, n14, 168
Godfrey-Smith, P. 2–3, 76–90 *passim*, 177 n1, 196 n2, 198 n6
Gold, T. 218
Goldman, A. I. 71
Goodale, M. 140
Goodman, N. 103
Gould, S. J. 179 n3
Grahek, N. 133, 135
Grave, S. A. 29, 32 n4

Hardin, C. L. 136, 139
Harman, G. 117 n6
Harré, R. 251
Hart, H. L. A. 40
Hawthorne, J. 216, 224
Hazen, A. P. 120 n11
Heil, J. 7, 65, 244–57 *passim*
Heller Illusion 113–15, 117–18
Hempel, C. 18 n10, 30, 82 n10
Hilbert, D. 118, 122 n12
Hobbes, T. 47
Hopes 202
Horgan, T. 137
Hume, D. 5, 40, 53, 54, 55, 60, 78, 200, 250
  Distinctness principle 5, 54, 58, 60, 78, 145, 150, 152, 154, 155–6
  on causation and laws 53, 54, 55, 56, 58, 69, 216, 217, 219, 251, 254
  motivation 201
  on perception 34
Husserl, E. 238–9

Identity of Indiscernibles 55, 57
Identity statements 4, 14, 42, 126–7, 163–4, 234
Identity theory 1, 2, 12–14, 24–7, 29–47 *passim*, 58, 67, 69, 76, 128, 163, 164, 231–2
  see also, Australian materialism, physicalism
Information 4, 8, 18, 97, 99, 101, 104, 108, 112–23 *passim*, 140, 173, 176–8, 188–93, 196–8, 203, 205, 230, 246
Individuation 35, 39, 42, 55, 57, 71, 103, 153, 245, 248, 254, 255
Induction, problem of 56
Inference to the best explanation 11, 20, 40, 147, 165, 171, 172
Intention, intending 78–9, 81 n7, 82, 166 n39, 176, 192, 196–8, 199 n9, 205, 246
  see also, purposive activity
Intentionality 6, 76, 100–1, 106–9, 126, 137, 138, 196–7, 206–8
Introspection 4, 5, 11, 17, 95, 101, 106, 107, 108, 109, 143–73 *passim*, 176, 177, 178, 181, 184–8, 189, 191, 193

Introspection (*cont.*)
  constitutive theory 143, 155–65
  infallibility of 4–5, 102, 143, 145, 147–50, 151, 159, 172
  self-scanning view 4, 5, 107–8, 109 n11, 143–73 *passim*, 188, 239

Jackson, A. C. 62
Jackson, F. 3, 7, 14, 19, 82 n11, 85 n17, 112–23 *passim*, 140, 145 n8, 150 n21, 152, 154 n26, 157 n29, 161 n34, 222, 252, 253
James, W. 89 n21

Kant, I. 40, 54–5, 60, 61, 106, 143, 144
Kim, J. 30, 249–50
Kind, A. 146 n14, 166 n40
Klein, C. 177 n1
Kneale, W. 237
Knowledge argument 3–4, 19, 85, 112, 116, 117 n6, 122–3, 140
Kriegel, U. 109 n11
Kripke, S. 126–7, 133, 229 n4

Lasnik, H. 186
Laws of nature 41, 57, 59, 69, 71, 72, 231, 250, 251, 254
LeDoux, J. 177 n1
Leibniz, G. W. 40, 57
Levine, J. 4, 97, 125–41 *passim*
Lewis, D. 1, 2, 4, 7, 9–17, 21, 33, 39 n13, 47, 81, 82, 84, 123, 164, 187 n11, 198 n6, 205 n17, 211, 246–8, 249
  and behaviourism 40–1, 44, 46
  counter-factual causation 24
  and the Identity theory 24–7, 29, 33, 40–6
  theoretical terms 43–6
Lewis, D. (Works)
  'An argument for the Identity theory' 1, 12, 21, 24–7, 42–3
  'How to define theoretical terms' 21, 43, 44
  'Mad pain and Martian pain' 14, 25, 27, 44
  'Psychophysical and theoretical identifications' 21, 43, 44, 82, 149 n18
Lewis, S. 24, 47 n20
Lewontin, R. C. 179 n3
Linsky, L. 195
Loar, B. 137, 207 n24
Locke, J. 59, 77, 78, 125, 128–32, 134, 138, 143, 144, 250
Logical Positivism 2, 31, 46
Lucretius 47
Lycan, W. G. 5, 6, 76, 79, 87 n19, 145 n6, 146 n9, 168 n41, 176, 181, 182, 195–208 *passim*
Lyons, W. E. 146 n10

McDaniel, K. 203, 205
McGinn, C. 191–2
MacIntyre, A. 71
Mackie, J. 55, 201 n11
Madden, E. H. 251
Malcolm, N. 181
Maley, C. J. 177 n1, 180 n6
Marcel, A. J. 109
Martin, A. 202
Martin, C. B. 12 n4, 18, 31, 32, 37, 65, 66, 71, 227 n2, 245, 251, 255
Masked priming 109
Materialism 1, 3, 6, 9, 10–12, 14, 18–21, 37, 38, 46, 47, 67, 76–90 *passim*, 92, 125, 126, 195, 211, 212, 215, 225, 233–4, 239, 245
  Australian 1, 2, 29–31, 32 n5, 33, 40–46, 47
  central-state theory of 4, 13, 25, 39–40, 43, 47, 67, 69, 71, 144, 145, 147–50, 159, 160, 162, 173
  see also, Identity theory, physicalism
Maudlin, T. 218–19
Maze, J. 55
Medlin, B. 9, 13, 14, 24 n1, 27 n3, 47, 67 n2
Mellor, D. H. 38 n12
Mental discourse 228, 240–1
Mental images 3, 11, 67, 94–5, 100–1, 229
Mental states 5, 7, 8, 21, 26, 27 n3, 29, 37–9, 44 n16, 45, 47, 67, 71, 77, 78–80, 92, 93, 94, 96 n4, 97, 99, 136, 143, 146–51, 153, 154, 159, 160, 163, 165, 168, 170, 172, 173, 176, 177, 179, 182, 183, 185–91, 196, 197, 198 n6, 199, 202, 204, 205, 207, 212, 214, 218, 222, 224, 228, 229
  as brain states 4, 13–15, 17, 25, 40, 41, 42, 43, 46, 77, 84, 126, 144, 145, 147, 164, 176, 195, 231, 253
  as causes 17, 38, 77, 92, 101, 133, 167, 197, 112–13, 225, 228, 233, 234, 240, 246, 247, 252, 254
  and consciousness 105–9, 118 n7, 161 n33, 181–3
  content of 6
  core/peripheral 2, 78–82, 83
  epistemic access to 26
  intentionality of 138
  as tokens 13, 43, 151 n23, 195
  as types 7, 13, 14, 92, 94–5, 98, 106, 165–6, 195, 196, 201, 214
  see also, introspection, multiple realizability thesis
Mercier, H. 180 n5
Millikan, R. 76
Mill, J. S. 57
Milner, A. D. 140
Montero, B. 85 n16

Moore, G. E. 11, 40, 117 n6
  Moorean common sense 234
Moran, R. 192–3
Multidimensional scaling 98 n5
Multiple realizability thesis 8, 78, 133, 134, 248–9
Murdoch, I. 24, 31, 40

Nagel, T. 71, 97, 195
Nemirow, L. 4, 123
Neutral monism 3, 85–90
Newton, Sir I. 250
Nichols, S. 146 n9

Ockham's (Occam's) razor 27, 31, 39, 41, 86, 87 n18, 232
  parsimony 3, 39, 41, 85–7
  see also, simplicity
Öğmen, H. 109
Oppenheim, P. 41
Orzack, S. H. 179 n3

Pain 4, 13, 14, 25, 27, 29–30, 35, 37, 42–6, 57, 62, 63, 65–6, 67, 72, 126–30, 132, 133–40, 157, 162, 163, 185–7, 201 n11, 234, 237 n8, 247–9, 255, 256–7
Panpsychism 85, 88
Papineau, D. 85 n16, 132
Pargetter, R. 14, 252
Particulars 40, 52, 53, 54, 57, 59, 65, 68, 71, 151, 154
  bare (haecceity) 57, 63
  see also, universals
Pautz, A. 97, 181
Peacocke, C. 117, 214 n3
Perception 2, 3, 4, 9, 34, 36, 37, 53, 60, 61, 62, 68 n3, 78, 82–3, 87, 92–109 *passim*, 112–16, 118–21, 128–9, 140, 144–5, 167, 168, 173, 176, 184, 188, 196, 206, 207–8, 229, 237, 241, 251
  direct realism 10, 33, 34, 52, 60, 61, 62, 67, 71
  higher-order 106–8, 109
  indeterminacy of 103–5
  phenomenology of 114, 115, 117, 119, 123
  of pressure 217
  representative theory 61, 62, 188
  see also belief, seemings
Perceptual monitoring 107, 145 n6
Perceptual discrimination 82, 83, 97–9, 100, 102, 104, 136, 137, 139, 207, 208
  Just noticeable differences 98–9, 102, 103
Persson, I. 203 n14
Phenomenalism 32, 35, 61, 62, 66, 78
Phenomenological fallacy 256
Phillips, B. 109 n11
Phlogiston 236
Physicalism 3, 17, 18 n10, 25, 44, 84, 85, 88, 132, 138, 214, 215, 220, 228, 232, 233, 247–9
  see also, Conceivability argument, knowledge argument, Materialism
Piccini, G. 177 n1, 180 n6
Place, U. T. 2, 9, 13, 17, 25–6, 29, 30–3, 39, 46, 125, 232, 245, 256
  'Is consciousness a brain process?' 25, 29, 31
Polger, T. W. 117 n1
Popper, Sir K. 86, 87 n18
Presley, C. S. 31 n2
Price, Huw 218, 227–8
Price, H. H. 34–5, 37
Prior, E. W. 14, 252
Properties 2, 3, 18, 19, 35 n9, 52, 54, 56, 57, 58, 59, 61, 62, 63, 66, 67, 70, 71, 88, 93–6, 98–105, 108, 112, 116–23, 127, 130, 132, 134, 136, 153, 154, 157, 181, 254
  categorical 7, 38, 39, 55, 57, 58, 64, 68–70, 72, 195, 204, 252–5
  characters 57, 58–9, 63, 65, 128, 132, 137, 139, 141
  complex 30–1
  epiphenomenal 138, 247, 253
  fundamental 84
  higher-order 247–9, 252
  intrinsic 3, 18, 59, 88, 109, 119
  mental 85, 93, 94, 97, 101, 102, 104, 105, 107, 108, 223, 228, 247–9, 250, 252, 253, 255
  phenomenal 31, 127, 132, 134, 136
  physical 3, 62, 66, 93, 94, 95, 96, 118, 128, 129, 211 n1, 218, 220, 228, 231
  see also dispositions, qualia, qualities, relations
Proprioception 144 n2, 145 n6
Psychosemantics 206, 207, 208 n25
Purposive activity 5, 107, 108, 109, 176, 177–80, 182–91, 196–208
Putnam, H. 8, 41, 44, 206, 229 n4

Qualia 30, 31, 43, 57, 140, 214, 222–5, 250
  see also, epiphenomenalism
Qualities 54, 55, 57, 63, 71, 137, 254–7
  mental 3, 94–106 *passim*
  primary 62, 125, 129–31
  quality space 98–101, 104, 105, 106, 136–9
  secondary 9, 35, 93, 125, 128–31
  sensory 129, 131, 144 n6
  tertiary 128, 130, 131
  see also, properties, qualia
Quine, W. V. O. 229 n4

Ramsey, F. 3, 29, 44, 45–6, 47, 78, 80–3, 196 n2
Ramsey sentence 8, 20, 43, 45, 82, 164, 198 n6
Ravenscroft, I. 214
Realism 33, 34, 58, 70, 208 n25, 253, 254

Realism (*cont.*)
  Australian 253
  critical 30
  scientific 31, 37
  see also, colour, perception
Relations 53, 54, 55, 59, 61, 71, 85, 89 n21, 120, 123, 139, 151, 155, 173, 184, 208, 218, 219, 249, 250, 254
Representation 6, 76, 83, 84, 99–100, 102, 106, 108, 113 n2, 116, 130, 138, 184
Revelation 19, 223
Robinson, H. 140
Robinson, Z. 177 n1, 180 n6
Rosenthal, D. 1, 3, 92–109 *passim*, 136, 137, 177 n1
Russell, B. 18, 89 n21
Ryder, Z. 182
Ryle, G. 15, 16–17, 24–6, 30, 31, 35, 38, 39, 40, 52, 58, 63–9, 71, 78, 79, 81, 195–6, 229, 230, 236, 245, 251, 252, 253
  The Concept of Mind 1, 15, 17, 24, 40, 52, 63, 65, 195, 251
  see also, behaviourism (Rylean) dispositions (Rylean)

Saito, M. 186
Sanford, D. 147 n15, 152, 156 n28, 162–3, 164 n28
Scheffler, I. 45–6
Schlick, M. 30, 232
Schueler, F. 199, 200, 201 n11
Science 18, 38, 41, 53, 77, 86, 119, 215, 228, 230, 232, 233, 234, 255
  authority of 9, 10, 11, 15, 19, 36–7, 231
  and philosophy 29, 30, 36, 227
Scott, R. B. 98, 104
Searle, J. 30
Secondary qualities, *see* qualities
Seemings 103, 104, 119, 237–9, 241
  see also, perception
Self-blindness 146, 165–8, 184–6, 188 n12, 191
Self-knowledge 4, 146, 167, 168
  see also, introspection
Sellars, W. 100 n6, 238–9
Sensations 3, 31, 94, 95, 100, 101, 127, 128, 136, 139, 140, 144 n6, 228, 229, 240
  bodily 10, 35, 37, 58, 62, 92 n1, 96 n4, 125, 128, 130, 135, 137, 256, 257
  see also, pain
Sense-data 32, 34–5, 60–1, 62, 117, 237
Shaffer, J. 40–1
Shoemaker, S. 4, 5, 143, 144, 146–7, 165–73, 177, 184–8, 191–3, 248
Sider, T. 227 n1
Siegel, S. 100 n6, 184 n9
Simplicity 86–7, 89 n21, 114
Smart, J. J. C. 2, 7, 9, 11, 13, 17, 24–7, 29–33, 37, 39–44, 46, 85, 125, 147, 150 n21, 211, 217, 219, 231–3, 237–8, 245, 256
  'Sensations and brain processes' 25, 32 n2, 237–8
Space and time 25, 53, 55, 56, 57, 59, 61, 63, 71, 130, 183, 217–18, 230
  spatio-temporal relations 40, 53, 54, 55, 57, 59, 71
Sperber, D. 180 n5
Squires, R. 12 n5, 70, 71
States of affairs 57, 151, 153–7, 192, 222
Sterelny, K. 77 n2, 206 n22
Stevenson, J. T. 67
Stich, S. P. 76, 146 n9, 206–8, 214
Stoljar, D. 5, 88, 154–5, 176–93 *passim*
Stove, D. 55, 77
Strawson, G. 223
Strawson, P. 40, 235
Supervenience 59, 88 n20, 121, 138, 166, 170–1, 217, 249–50, 253
Sydney Free Thought Society 55

Tanney, J. 31
Taylor, R. 55
Theoretical terms 2, 21, 29, 43–7
  see also, Lewis, Ramsey, Ramsey sentence
Thomasson, A. L. 6, 227–42 *passim*
Thornton, J. B. 55
Tienson, J. 137
Tooley, M. 219
Topic-neutral account of sensation-reports 25, 26, 29, 41, 43, 217, 232
Transitivity of identity 102, 103
Truthmakers 39, 204, 229 n3, 248
Tye, M. 113 n2

Universals 35 n9, 37, 40, 52, 56, 57, 63, 71
Urmson, J. O. 71

Van Fraassen, B. 205 n17
Velleman, D. 120 n11

Walters, R. 55
Weiskrantz, L. 109
Will, the 17, 78, 196, 199, 208
Williams, D. C. 24, 25, 39 n13, 42
Williamson, T. 154 n25
Wilson, N. L. 42
Wisdom, J. 32, 40
Wishes 5, 95 n3, 202
Wittgenstein, L. 31, 40, 78, 229, 245, 246

Zimmerman, A. Z. 146 n10
Zombie arguments 6, 85, 211–25 *passim*
  see also, conceivability argument